DOCUMENT RESUME

ED 227 592	EA 015 517
AUTHOR	Murnane, Richard J.; Pauly, Edward W.
TITLE	How the Financing of the Public Schools Affects Their Ability to Educate. Final Report of Research Project.
INSTITUTION	Yale Univ., New Haven, Conn. Inst. for Social and Policy Studies.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Aug 82
GRANT	NIE-G-79-0084
NOTE	383p.
PUB TYPE	Reports - Research/Technical (143) -- Viewpoints (120) -- Information Analyses (070)
EDRS PRICE	MF01/PC16 Plus Postage.
DESCRIPTORS	Academic Achievement; Behavioral Science Research; Comparative Analysis; Educational Environment; *Educational Finance; Educational Policy; Elementary Secondary Education; Equal Education; Input Output Analysis; Outcomes of Education; *Politics of Education; Private Schools; *Productivity; Public Schools; *Research Problems; *School Effectiveness; Seniority; Student Teacher Relationship; Tax Credits; Teacher Supply and Demand; Tuition
IDENTIFIERS	*Public and Private Schools (Coleman et al)

ABSTRACT

This research report is comprised of a series of 14 papers, each accompanied by a bibliography. Murnane's nine self-contained articles are grouped into four projects. The first consists of three papers that interpret the results of quantitative research on school effectiveness in a manner that highlights the active behaviors of participants and demonstrates that school policies influence school outcomes through the behavioral responses of participants. The second project applies the behavioral response idea in exploring why it has been difficult to build a powerful research program in education, and why seniority rules for teachers are more productive than many analysts have believed. The third project applies the behavioral response idea to the debate over the relative quality of the education provided by public and private schools. The fourth project consists of an empirical study establishing that there is a selective teacher attrition, during the first years on the job, of those perceived to be unproductive. Pauly's five essays provide the basis for a considerably revised conception of how teaching and learning are embedded in a system of school policies, institutions, and independent behavior. (MLF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227592

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Final Report of Research Project

HOW THE FINANCING OF THE PUBLIC SCHOOLS AFFECTS THEIR ABILITY TO EDUCATE

Funded by Grant NIE-G-79-0084

Submitted by
Richard J. Murnane
Edward W. Pauly
Institution for Social and Policy Studies
Yale University
New Haven, Connecticut 06520

August, 1982

EA 015 517

I. Introduction I-1 to I-6

II. Interpreting the Evidence on School Effectiveness II-1 to II-32

III. Making Sense of Research on School Effectiveness III-1 to III-9

IV. Input-Output Research in Education: Accomplishments, Limitations, and Lessons IV-1 to IV-23

V. Input-Output Relations and Innovation in Education V-1 to V-48

VI. Seniority Rules and Educational Productivity: Understanding the Consequences of a Mandate for Equality VI-1 to VI-38

VII. Evidence, Analysis, and Unanswered Questions: Coleman's New Study, Public and Private Schools VII-1 to VII-16

VIII. The Uncertain Consequences of Tuition Tax Credits: An Analysis of Student Achievement and Economic Incentives VIII-1 to VIII-33

IX. Comparing Public and Private Schools: The Puzzling Role of Selectivity Bias IX-1 to IX-35

X. Selection and Survival in the Teacher Labor Market X-1 to X-18

XI. Teachers Control Students, Students Control Teachers XI-1 to XI-15

XII. On the Political Nature of Classroom Teaching and Learning XII-1 to XII-12

XIII. How People in Schools Coordinate Themselves Without Intending To (And Other Features of School Life) XIII-1 to XIII-16

XIV. What Difference Do Classroom Interactions Make? Teachers, Students and Reciprocal Sovereignty XIV-1 to XIV-18

XV. Analysis When Conventional Analysis Won't Work: School Officials and the Uses of Interactive Analysis XV-1 to XV-25

During our research over the past three years, we continually searched for a powerful perspective from which to view the central question that motivated our work and which provided the title for our original grant proposal--"How the Financing of the Public Schools Affects Their Ability to Educate." The perspective which we developed, which runs through all of the products of this research, can be described as follows:

> The key to understanding how the financing of the public schools affects their ability to educate lies in the relationship between finances and the individual actions of teachers, administrators, students and parents. The people most directly involved in public schooling--teachers, pupils, and officials--often behave in quite independent ways, frequently resulting in unexpected policy outcomes. This is significant in two related ways:
>
> -To understand the effects of education policies (including fiscal policies) on achievement, we must understand how officials, teachers and students make choices in response to streams of policy; that is, the independently-chosen behaviors that link finances to outcomes must be explicitly addressed.
>
> -Since learning and school achievement are accomplished by the choices and actions of people in classrooms, it is necessary to consider the ways that classroom actions that are directly responsible for learning can be supported by extra-classroom decisions on finances and related matters.

Each investigator on this project developed and applied this perspective in a series of papers, all of which are included in this report. While the papers are separately authored, they all represent the results of almost daily interactions over a three year period.

Murnane's papers can be divided into four projects. The first consists of three papers that interpret the results of quantitative research on school effectiveness in a manner that highlights the active behaviors of participants in the schooling process and demonstrates that school policies influence school outcomes through the behavioral responses of students, teachers, administrators, and parents. These three papers are entitled:

---"Interpreting the Evidence on School Effectiveness," Teachers College Record, Fall 1981. This paper will also be included in the 1983 Yearbook of the American Educational Finance Association

---"Making Sense of Research on School Effectiveness: The Primacy of Human Resources in Schooling," Impact on Instructional Improvement, Summer 1981

---"Input-Output Research in Education: Accomplishments, Limitations, and Lessons," New Directions for Testing and Measurement, forthcoming.

The second project applies the behavioral response idea, exploring how this perspective can help us to understand a) why it has been difficult to build a powerful research program in education despite considerable federal attention to this goal; and b) why seniority rules for teachers are more productive in education than many analysts have believed. These papers are entitled:

—"Input-Output Relations and Innovation in Education" (with Richard Nelson)

—"Seniority Rules and Educational Productivity: Understanding the Consequences of a Mandate for Equality," American Journal of Education, November 1981

The third project applies the behavioral response idea to the debate over the relative quality of the education provided by public and private schools in the United States. A theme of these papers is that mechanistic comparisons of the relative quality of public and private education made without concern for the consequences of the active behaviors of families choosing schools for their children are not informative. The papers in this group are entitled:

—"Evidence, Analysis and Unanswered Questions," Harvard Educational Review, November 1981

—"The Uncertain Consequences of Tuition Tax Credits: An Analysis of Student Achievement and Economic Incentives." This will be published in a volume on tuition tax credits prepared by the Stanford University Institute for Research on Educational Finance and Governance and published by Temple University Press.

—Comparing Public and Private Schools: The Puzzling Role of Selectivity Bias (with Stuart Newstead and Randall Olsen)

The fourth project consists of an empirical study investigating a behavioral response of teachers. It takes up the question of whether particularly effective teachers or particularly ineffective teachers were more likely to leave an urban school district after one or two years of teaching than other teachers were. This project, which was funded in part by the NIE grant and in part by a grant from the Spencer Foundation, supports the main theme of Pauly and Murnane's research in that it is based on the idea that the active career decisions of teachers influence the quality of the teaching staffs of urban schools. The results of this research are reported in a paper entitled:

---"Selection and Survival in the Teacher Labor Market"

Murnane's original plan for disseminating the results of his research was to write a book. However, as the work came to address topics of interest to quite different groups, it seemed more appropriate to publish the results as a series of articles, directed to different audiences. For this reason, the research results were packaged as nine self-contained articles. Six of these have been published, or have been accepted for publication, in journals or in edited volumes written for policymakers. The other three articles are currently under review at scholarly journals.

Pauly chose a research strategy somewhat different from Murnane's, in that he decided to produce a series of five essays that form the core of a tightly integrated book. The first of these essays, "Teachers Control Students, Students Control Teachers," presents a model of the relationships among teachers and students in classrooms that stresses the active behavior of classroom participants and their mutual dependence on each other.

The second essay, "On the Political Nature of Classroom Teaching and Learning," explains how the interdependence of students and teachers in each classroom produces a political structure, which in turn impinges directly on teaching and learning behaviors in the classroom. Building on the first essay, he argues that the distinctive political arrangements made by teachers and students in a classroom have a powerful influence on student achievement.

The third essay, "How People in Schools Coordinate Themselves Without Intending To," explains how the school environment is shaped by the interaction of teachers and students in different classrooms within the same school.

The fourth essay, "What Difference Do Classroom Interactions Make: Teachers, Students and Reciprocal Sovereignty," explains why it is so critical to understand the interactions of people in classrooms and schools—namely, that policies can only affect school outcomes through their impact on the complex and robustly-established interactions of school participants.

The fifth essay, "Analysis When Conventional Analysis Won't Work: School Officials and the Uses of Interactive Analysis," explains why conventional analysis of school problems and solutions is not effective and suggests that school officials, often without knowing it, engage in a type of analysis different from that taught in administration courses. This type of interactive analysis pays attention to the mutual interdependence of people in schools.

Taken together, these essays provide the basis for a considerably revised conception of how teaching and learning are embedded in a system of school policies, institutions, and independent behavior. This new approach, to be laid out in a forthcoming book based on these five essays, has significant implications for policy analysis in general and for education policy in particular.

In summary, we believe that the products of this research project develop in a variety of ways and in a variety of contexts the following theme:

> It is crucial for policymakers to appreciate that it is people in classrooms who will make the choice and do the work of learning; only policies that consider, support and take advantage of strongly motivated classroom behavior, rather than try to replace or control that behavior, will be effective. Policies can be assessed in terms of how useful they can be to teachers, students and officials, and how supportive are the mechanisms and settings created by policy decisions for the use of people in schools and classrooms.

Interpreting the Evidence on School Effectiveness

by

Richard J. Murnane

Institution for Social and Policy Studies
Yale University

December 1, 1980

This paper is based on research supported by grant NIE-G-79-0084 from the National Institute of Education.

The author would like to thank David Cohen, Michael Krashinsky, Charles Lindblom, Richard Nelson and Edward Pauly for helpful comments on an earlier draft.

ABSTRACT

This essay provides an interpretation of school effectiveness research that explains puzzles in the empirical findings and clarifies what this research can and cannot tell us. Section I reviews and analyzes the quantitative studies of school effectiveness. The main theme is that the primary resources of schooling are the human resources, teachers and students. Physical facilities, class size, curricula, and instructional techniques can best be understood as secondary resources that affect student learning through their influence on the behavior of teachers and students.

Section II explains why such research will not provide reliable information about the effects on student achievement of policies designed to improve the school resources available to children. What is needed for effective policymaking, and what current quantitative research does not capture, is information about the behavioral responses of teachers, students, and families to changes in resource allocation mechanisms.

Section III discusses research questions that do focus on the behavioral responses of teachers, students and families. The essay concludes with a brief analysis of the importance of decisionmaking processes. The argument is presented that teachers' unions and other interest groups can play a positive role in decision making by providing information about critical behavioral responses.

The purpose of this paper is to examine what has been learned from quantitative studies of school effectiveness and to assess the implications of the research results for public policy. Now is a particularly appropriate time to discuss this research because in these days of declining enrollments, severe budget constraints and court mandated school finance reform, the results of studies in this tradition are often cited in public policy debates concerning the role of public schools. These public policy debates frequently center on questions such as:

- Are there systematic differences in the quality of education provided in public schools?
- What school resources really make a difference?
- What public policies should be implemented to improve the quality of education provided to disadvantaged children?

One of the goals of this paper is to explain the contributions that research in this tradition has made in providing answers to these questions and to clarify what this research can and cannot tell us.

Section I of this paper presents a critical review of the results of quantitative studies of school effectiveness. Section II explains the limits of this type of research. In particular, this section points out why such research cannot provide reliable information about the effects on student achievement of policies designed to improve the school resources available to children. The crux of the message is that this type of research does not provide information about the behavioral responses of teachers, students, and families to changes in resource allocation mechanisms. Section III discusses strategies for taking into account the behavioral responses of the key actors in the educational process in formulating public policy.

I. QUANTITATIVE STUDIES OF SCHOOL EFFECTIVENESS

In the last fifteen years, a large number of quantitative studies of the relationship between school resources and student achievement have been conducted. Some are called input-output studies, others, educational production function studies, and others simply multivariate studies of school effectiveness. Definitions of school resources have differed, as have the measures of student achievement. Despite these differences, these studies, which we shall call simply quantitative studies of school effectiveness, share a basic methodology and can be viewed as examples of a particular research approach. In this approach, no attempt is made to manipulate experimentally the school resources that children receive. Instead, it is "natural experiments"--the variation in school resources created by the operation of a school system--that provide the data base for analysis. In essence, the research strategy can be viewed as taking a snapshot of a school system at work. The key parts of the snapshot are information on the school resources that children receive at a point in time and one or more measures of student progress. Sometimes the snapshot also includes information about students' family backgrounds. Multiple regression techniques are used to estimate the impact of individual school resources on student achievement.[1/]

In the last fifteen years we have learned a great deal about how to take more accurate snapshots of schools at work. In particular, we have learned the importance of using the individual child as the unit of observation, of using children's progress as the measure of school effectiveness (instead of the student's achievement level), and of identifying the school resources that each child actually receives (rather than using the average resources present in the school or the school district). In addition, the definition of school resources has become much broader and more sophisticated. The first studies focused on physical facilities, library books, student-teacher ratios and school size. In recent studies, the definition of resources has been expanded to include characteristics of teachers and classmates, indicators of teacher quality, the amount of time devoted to learning tasks, and descriptions of instructional techniques. These improvements in methodology have increased the ability of research in this tradition to provide reliable information about the impact of school resources on student achievement in the particular times and places that are studied.

What have we learned from quantitative studies of school effectiveness? The most notable finding is that there are significant differences in the amount of learning taking place in different schools and in different classrooms within the same school, even among inner city schools, and even after taking into account the skills and backgrounds that children bring to school. The importance of this result, found in all four studies which have addressed this question, cannot be underestimated (Armor, et al., 1976; Hanushek, 1971: Murnane, 1975; Murnane and Phillips, 1979). It provides clear support for the belief of most Americans—that schools matter. It also provides support for the position that it is worthwhile devoting attention to the question of why some schools provide better education than other schools do, despite our limited success in answering this question.

Having determined that more learning takes place in some schools and classrooms than in others, researchers turned to the question of whether the differences can be explained by differences in school resources. There is no unequivocal consensus regarding the role of any school resource in contributing to student achievement. However, a judicious interpretation of the evidence (including the research methodology as well as the pattern of coefficient estimates) does suggest some tentative conclusions.

Before turning to discussions of individual resources, it is important to note that in all of the studies discussed in this essay, student achievement is measured by improvements in scores on standardized tests of cognitive skills. These tests are by no means problem free (for example, see Averch, 1972). However, they do provide the best available measures of student achievement that can be used in large scale studies.[2/]

Teachers

To most Americans, quality of education is synonymous with quality of teaching. Thus, it is not surprising that the role of the teacher has been a central focus of quantitative research on school effectiveness. The research strategy used to study teachers has been to include measurements of teacher characteristics in the vector of school resources that is related to student achievement. The choice of the teacher characteristics included in any study has depended primarily on the availability of data. Thus, it is often difficult to compare results across studies. Despite this problem, however, the results have been informative.

Virtually every study of school effectiveness finds that some attributes of teachers are significantly related to student achievement, and certain attributes play an important role in several studies. In particular, the intellectual skills of a teacher, as measured by a verbal ability test (Hanushek, 1971; Hanushek, 1972) or the quality of the college the teacher attended (Summers and Wolfe, 1977; Winkler, 1975) tend to be significant. Teachers with some experience are more effective than teachers with no experience (Hanushek, 1972; Murnane, 1975; Murnane and Phillips, 1978)--although one study reports a significant exception to this conclusion (Summers and Wolfe, 1977).[3/] Teachers with high expectations for their students are effective in helping children to acquire cognitive skills (Link and Ratledge, 1979). Recent studies in which large samples were examined indicate that there are significant interaction effects between the characteristics of teachers and students (Kiesling, 1979; Summers and Wolfe, 1977). In other words, some teachers are more effective with certain types of students than with other types of students.

One interesting negative result present in many studies is that teachers with Master's Degrees are no more effective on average than teachers with only Bachelor's Degrees. At the same time, studies have found that teachers who voluntarily attended post-graduate courses are particularly effective (Hanushek, 1971). This suggests that voluntary participation in post-graduate education may be a signal of high motivation--an attribute that is difficult to measure, but which administrators feel is crucial to a teacher's effectiveness. It may be that when the pay increment for possession of a Master's Degree was first introduced into teachers' salary schedules, it was justified by productivity differences. At that time, only a small percentage of teachers had Master's Degrees, and these may well have been the most highly motivated teachers. Today, however, when a majority of teachers have advanced degrees, and when some states require that teachers obtain MAs to earn permanent positions, the degree is no longer a signal of a particularly high level of motivation.

One final result concerning teachers is that supervisors know in general who the more effective teachers are. Two studies (Armor, et al., 1976; Murnane, 1975) have analyzed the relationship between principals' evaluations of teachers and the effectiveness of the teachers as measured by their students' progress on standardized tests. In both studies, the evaluations were significantly related to student test score gains (and there is evidence that the evaluations were not based on the test results).

Peer Groups

The school related research on peer groups asks whether a child's achievement (or attitudes) is affected by the characteristics of the children with whom he/she interacts in school. This is an extremely important public policy question since peer groups are a resource that cannot be equalized by simply providing more dollars to schools serving needy children. If peer groups are critical, as Coleman suggested in his 1966 Report, the meaning of equality of opportunity must be reconceived.

Two problems have hindered research on peer group effects in schools. The first problem is the difficulty in identifying the "peer group". In practice, the characteristics of individual data bases determine whether a child's peer group is defined as the other children in the classroom, in the grade level, or in the school as a whole. Whether a particular definition provides accurate information about the children with whom a child actually interacts depends on the organization of the school--in particular, on the extent to which self-contained classrooms, tracking and homogeneous grouping are used. Only rarely have studies even attempted to control for grouping practices.

The second problem concerns the attributes of peers. Most parents want their children to interact with other children who share their values and are motivated to succeed in school. However, these noncognitive traits are very difficult to measure. As a result, in most studies peers are characterized by race, achievement, or family income. Differences in results across studies may be due to the fact that in different samples, the observed characteristics of peers are differentially related to the unobserved values and attitudes. The significance of this problem for public policy is discussed in Section II.

Despite these problems peer group research has begun to reveal some patterns. In particular, there is evidence that elementary school children with low initial skill levels who attend schools in which the average achievement level is relatively high make more progress than such children who attend schools in which the average achievement level is relatively low (Henderson, et al., 1978; Summers and Wolfe, 1977). There is similar evidence regarding socioeconomic status. Elementary school children from low SES families who attend schools with a high proportion of high SES students make more progress than children who attend schools in which most children come from low SES families (Winkler, 1975).

The evidence in regard to racial composition is more difficult to interpret. Some evidence suggests that both black and white students who attend schools in which the racial composition is in the 40-60% range make more progress than students in schools that are more segregated by race (Summers and Wolfe, 1977). Other evidence suggests that racial composition does not matter to either white or black students until the proportion of black students becomes quite high. Above a critical level (perhaps different for black and white students) achievement is decreased as the proportion of black students increases (Hanushek, 1972). Still other evidence indicates that black students who once attended racially segregated elementary schools subsequently do less well in racially mixed junior high schools than in segregated schools (Winkler, 1975).

The explanation for the differences in the racial composition findings may lie in the extent to which racially desegregated schools were in fact "integrated" in the sense that students felt comfortable and communicated with with each other. For example, black children who moved from a segregated elementary school to a racially mixed junior high school may have encountered a desegregated but not truly integrated environment. The unfamiliar confrontation with many better prepared white students may have been a threatening, discouraging experience that led to lower achievement.

A final peer group issue concerns the effect of student body composition on the achievement of "advantaged" children. Summers and Wolfe (1977) found that the progress of children with high initial test scores was not subsequently affected by the ability distribution of the children in their schools. Henderson, et al. (1978) found that children with high initial test scores gained just as much from being in classes in which the average achievement level was high as children with low initial test scores did. However, the effect on individual achievement of improvements in average class achievement was greater at the low end of the average achievement distribution than at the high end. The authors interpret this result as indicating that a policy of redistributing students in order to equalize the average achievement in every class would lead to large increases in the achievement of children in "slow" classes and small decreases in the achievement of children in "fast" classes.[4/]

Thus, it appears that children disadvantaged by low initial achievement or low SES benefited from attending schools with more fortunate students, while the cost to the more fortunate students in these schools in terms of decreased achievement was small. As we discuss later, however, the definition of a small cost lies in the eye of the beholder; if parents feel that mixing the race, ability, or SES of students reduces the quality of education for their child, they may respond in a manner that defeats the policy.

Class Size

The impact of class size on student achievement is perhaps the most thoroughly researched question in education. The reason is that class size is a highly visible indicator of quality to many parents and teachers; it is also a good indicator of per pupil instructional costs since teachers' salaries comprise the bulk of instructional expenditures. Consequently, the class size issue is of great interest to both advocates of better education and proponents of tax relief. Despite the extraordinary volume of research, there is no consensus on the role of class size. Evidence exists to support both smaller classes and smaller budgets. A recent synthesis of past research by Glass and Smith (1978) found that average student achievement was much higher in very small classes than in classes with twenty or more students. However, average achievement in classes with twenty students was only marginally higher than average achievement in classes with thirty or forty students. This does not offer much consolation to educators in urban areas concerned with increases from 28 students per class to 30 students.

Why is the role of class size so elusive? There are two parts to the answer to this question--both concern limitations in the ability of existing research to capture salient aspects of the education process. The first problem is that the effect of class size surely depends on a teacher's instructional strategy. Class size would matter less in a class in which the teacher provided instruction to the entire class simultaneously than in a class in which the teacher relied heavily on individualized instruction. In principle, this interaction effect between class size and instructional strategy can be investigated using multiple regression if the sample size is sufficiently large. In practice, however, this is very difficult to do because reliable information on instructional strategy can only be obtained using expensive observational techniques. As a result, studies using such techniques usually employ very small samples.

A second and related problem concerns the insensitivity of existing research strategy to the effects of class size on the children most affected by this variable. It seems plausible that the cost of a large class may not be borne proportionately by all of the students in the class. Instead, the cost is borne primarily by children with learning problems who do not profit from instruction geared to the average achievement level in the class. In a small class the teacher may be able to find the time to provide particular attention to such children. It is frequently not possible to examine this hypothesis effectively because children with special learning problems tend to be absent from school more often than other children (Murnane, 1975). As a result, they are very likely to miss at least one of the two standardized tests that provide the measure of student progress. Consequently, children of this type have a disproportionately high probability of being excluded from samples used in school effectiveness studies.

Instructional Time

Recently, attention has focused on classroom time as a school resource. Interest in the role of time stems from the fact that school policies concerned with the length of the school day, the school year, and the number of subjects that are studied all affect the amount of time available for work on basic skill development. The first results on the role of time are encouraging, in that several studies report systematic relationships between measures of time use and student learning. However, at this point it is difficult to interpret the results because the analyses have used three different definitions of time. The first definition is the amount of time children spend in school (Wiley and Harnischfeger, 1974). The second is the amount of time devoted to basic skill development (Kiesling, et al., 1979). The third is the amount of time children spend "on task," actually working at basic skill development (Bloom, 1974; Thomas, 1977). Clearly, the third definition is the most relevant to learning basic skills. However, time on task is not a policy variable, and its relation to the definitions of time that can be manipulated by policy depends on the behaviors of students and teachers in ways that are not understood. The value of research on the role of time in improving education will depend on the success of efforts to understand how teachers and students transform aspects of time that are subject to public policy into the amount of time students spend "on task."

Physical Facilities

Physical facilities--for example, the number of library books in the school, the quality of the science labs, the size and age of the school--played a prominent role in early school effectiveness research. The reason for this interest was that physical facilities were the capital in the production process, and capital plays a central role in the economic models from which this research stemmed. However, the early studies did not find these indicators of capital to be systematically related to student achievement. (Moreover, as evidence began to accumulate concerning the importance of human capital, attention shifted to developing better measures of human resources.) Thus, the current conclusion is that the physical resources available in a school in a particular year are not systematically related to the achievement of the students in that year.

Does this mean that physical facilities do not matter? Perhaps. However, an alternative interpretation is that the quality of the facilities influences which teachers and children attend a particular school. This mechanism is not captured in the snapshot methodology used in quantitative studies of school effectiveness. We will develop this argument in greater detail in Section II.

Instructional Strategies and Curriculum

Instructional strategies and curriculum have long been the focus of a great deal of educational research. The primary reason is that research evidence indicating that particular instructional strategies or curricula were clearly better than alternatives would have direct implications for policy. Schools could purchase new curriculum packages. Colleges could train aspiring teachers in the use of the most successful instructional techniques.

Unfortunately, despite a great many studies and countless publications, no unequivocally superior curricula or instructional strategies have been found. Many studies report that students achieved at an exceptionally rapid rate when taught with a particular curriculum or instructional strategy. However, time after time, these successes have not been replicated in other sites, or even maintained in the original sites over a long period of time.

The most compelling explanation for the inability to replicate successes is that the same curricula and instructional strategies are used in very different ways in different sites. For example, Chall, in her well known book, Learning to Read: The Great Debate (1967), points out that even the basic distinction between the phonics approach to reading and the sight reading approach is not clear-cut when one observes their use in a number of classrooms. Similarly, Van Deusen Lukas (1975) reports enormous variation in the actual educational practices taking place in classrooms using the same innovative instructional approach.

Developers of innovative curricula or instructional strategies often interpret these findings as evidence that the problem lies in the lack of fidelity to the technical characteristics of the particular curriculum or instructional technique. Implicit in this view is the assumption that teaching and learning can be viewed as a stable, well defined production process, similar to growing hybrid corn. Fidelity to the details of the superior technology is thought to be possible and to result in increased productivity.

An alternative response to the evidence on the variation in practice is that such variation is unavoidable and in fact is crucial to effective teaching. A necessary condition for effective teaching may be that teachers adapt instructional strategies and curricula to their own skills and personalities, and to the skills, backgrounds and personalities of their students. In this view of teaching and learning, the technical characteristics of instructional strategies and curricula are not, by themselves, the critical components. Instead, what matters is the extent to which teachers are willing and able to adapt the curricula or instructional strategy to their needs and to the needs of their students (Berman and McLaughlin, 1978).

Summary of School Effectiveness Results: Primary Resources and Secondary Resources

We have learned a great deal from quantitative research on the determinants of school effectiveness. The most important lesson is that schools make a difference. Even in inner cities in which virtually all of the children attending public schools come from relatively poor families, there are important differences in the amount of student learning taking place in different schools and even among classrooms in the same school. A second lesson is that teachers are a critical resource. Children learn more when they are taught by talented, highly motivated teachers who believe that their students can learn and who structure the school day so that students spend large amounts of time "on task," working at basic skill development. We have learned a little about how to identify such teachers. However, it also appears that no set of observable characteristics provides a reliable composite picture of the effective teacher.

The research results also indicate that the composition of the student body matters. In the natural experiments that have been studied, disadvantaged children who attended schools which served a significant number of children from more advantaged backgrounds profited from this experience.

Quantitative research on school effectiveness began with a broadly specified input-output model that was agnostic on the role played by particular school resources. In the model, a large number of resources were treated in parallel fashion. A critical survey of this research indicates that the primary resources are teachers and students. It is on these human resources that researchers should concentrate, since they are poorly understood, play a central role in policy choices, and appear to dominate other resources.[5/]

Physical facilities, class size, curricula, and instructional strategies can be seen as secondary resources that affect student learning through their influence on the behavior of teachers and students. This perspective has two significant implications. First, current research methodology which employs a snapshot approach to examine the impact of school resources on student achievement may be inappropriate for measuring the influence of secondary resources. For example, it may be that these resources affect student achievement by influencing which teachers and children are found in particular schools. This mechanism, which is described more fully in Section II, is not captured with the snapshot methodology. The second, and related, implication is that research on the role of these secondary resources should concentrate on their impact on the behavior of teachers and students. We will return to this theme in Section III.

II. THE POLICY PROBLEM

In a nutshell, the policy problem is how to design policies that will provide more children with the school resources that contribute to rapid learning. Part of the difficulty in fulfilling this task stems from our limited understanding of what these resources are. However, research results provide increasing guidance concerning the resource configurations that are associated with high rates of student learning in ongoing educational systems.

A greater difficulty stems from the fact that resource configurations in ongoing systems result from a large number of institutional mechanisms, internal labor market rules and customs, and from the responses of teachers and students and families to these mechanisms. For example, the allocation of teachers to schools is determined by seniority rules and the decisions of the more senior teachers. Which children attend particular schools is determined by rules concerning attendance boundaries, and by family location decisions. The relationships between resources and student achievement that are observed in the natural experiment research are conditional on the resource configurations present in the school system. The process which created these resource configurations is not considered in the analysis.

To change the resource configurations in a systematic way requires altering one or more of the formal or informal institutional mechanisms. Any alterations in the institutional mechanisms will elicit behavioral responses on the part of teachers or pupils and their families. These behavioral responses may well alter the very relationships just surveyed between observable inputs and student learning.

Some readers may believe that the preceding paragraph simply reflects the excuses of a timid researcher afraid to pursue the policy implications of his work. They may point out that in the substantive area from which this research tradition stems--production of hybrid corn--rules of thumb were also used by tradition-bound farmers in determining combinations of seed, fertilizer, and other inputs. Yet there is clear evidence that convincing these farmers to abandon their rules of thumb and instead to allocate resources in the proportions indicated by the research findings resulted in significant increases in their productivity. Why is education so different?

The key difference is that in corn production, the key inputs, seed, water and fertilizer, are inanimate and their productivity depends only on the resource mix and on the weather, not on the method by which the resource allocation is determined. In education, the key resources are students and teachers, whose behavior and productivity are very sensitive to the methods used to allocate resources. This does not mean that policies cannot be altered. However, it does mean that effective policy analysis must take into account the behavioral responses that changes in resource allocation mechanisms will elicit.

Two examples may help to clarify the role of behavioral responses. The first concerns policies designed to take advantage of peer group effects. Recall that research has indicated that low SES children who attend schools with more affluent peers make more academic progress than poor children who attend schools with uniformly poor students. This has led to a number of policies designed to increase the mixing of students by class, ability or race. There has been enormous variation in the success of these policies. However, in a significant number of cases, the anticipated beneficial results have not been realized.

The reason may be that the middle class children who attend integrated neighborhood schools voluntarily as a result of their parents' decision to live in an integrated neighborhood may be different in unobserved critical ways from middle class children who attend schools that are desegregated as a result of a conscious policy such as court ordered busing. In particular, parents choosing to live in integrated neighborhoods and to send their children to public schools reveal by their choices the belief that public schools can provide their children with an adequate education. This belief, coupled with parental support and positive attitudes toward the other children in the school, may be critical in making the school a place where all children can learn. It may be for this reason that the research results indicate that poor children who attend such schools learn more than poor children who attend schools segregated by class or race.

Parents choosing to live in middle class enclaves may not share these attitudes toward urban public education and towards children from poor families. Without these critical, but unobserved attitudes, the policy of mixing children from different classes may not result in high quality education.

The second example concerns declining enrollments and teacher layoffs. Many school districts, faced with declining student enrollments and severe fiscal constraints, are forced to lay off a significant number of teachers. In most districts the layoffs are determined by seniority rules. However, some administrators have argued that this is inefficient since under this system many effective teachers are laid off while less effective, but more senior (and more expensive) teachers are retained. In some districts, administrators have dictated that those teachers who are designated by their principals as less effective will be laid off. Advocates of this policy point to the research evidence indicating that teachers do differ significantly in their effectiveness and that the evaluations of administrators do reflect teacher performance.

There is very little systematic evidence concerning how either layoff policy has affected the quality of education provided to children. However, there is limited evidence, much of it anecdotal, that the latter policy has been less successful in some districts than was hoped, for several reasons.

First, effective teachers may resign, not because they anticipate losing their positions, but rather because they find that the competitiveness bred by this system diminishes the enjoyment that they derive from their job (Jackson, 1968, pp. 119-135). Second, the quality of education provided in schools in these districts may decline as teachers adjust their behavior to take into account the fact that they are being compared with their colleagues. This can take the form of reluctance to share teaching materials or to help a fellow teacher deal with a particularly difficult child. Third, over time, as teachers alter their behavior, principals may find that their evaluations of teachers no longer reflect performance as well as they once did.[6/] (The studies that found that principals' evaluations accurately reflect teacher performance were carried out in districts where this information was not used in layoff decisions; consequently, the evaluations did not evoke the behavioral responses just described.)

The point of these two examples is to illustrate the types of behavioral responses that policies designed to alter resource allocations can elicit. In some cases the behavioral responses are obvious--for example, when middle class families withdraw their children from public schools rather than have them participate in a busing program. In other cases, the results may be more subtle. For example, in terms of socioeconomic status and other observable indicators, parents whose children are bused to desegregated schools may appear identical to children living in urban areas and attending neighborhood schools with many poor children. However, in unobserved dimensions, such as attitudes, the parents may be quite different, and the schools may be made different by contrasting levels of parental support.

The point of this section is not to argue that nothing can be done. There is a wide range of policies that can be used to alter resource allocations. Each of these will elicit a behavioral response, but the responses will differ. For example, the creation of magnet schools is an alternative to busing for promoting school desegregation. Unlike busing, magnet schools may evoke the positive parental support that is important to successful schooling. Early retirement programs are an alternative to layoffs for reducing the size of the teaching staff. These programs may permit the retention of talented young teachers without evoking the dysfunctional behavior that may accompany layoffs based on merit. The central point is that policy planning must take into account the behavioral responses that policies designed to alter resource allocations will elicit.

III. PRIMARY RESOURCES AND BEHAVIORAL RESPONSES: NEW QUESTIONS FOR RESEARCHERS AND POLICYMAKERS

This essay emphasizes the importance to student achievement of behavioral responses by teachers and students, the primary resources of schooling. These responses to institutional rules, and to the quantity and quality of secondary resources, determine first of all which children and teachers will participate in public schooling. They also influence the attitudes, expectations and motivations of the participants and ultimately the quality of the learning environment in particular schools and classrooms. Given the importance of these behavioral responses, it seems important to learn more about them. The following are a sample of research questions motivated by the behavioral response perspective:

- What factors influence teachers' participation in, and departure from, schooling as a career? In particular, under the existing system of compensation for teachers which rewards longevity and degrees, are effective teachers more likely to leave public school teaching than ineffective teachers are?
- Are particular working conditions critical determinants of teachers' decisions to leave public school teaching?
- Does class size influence the way teachers allocate classroom time among students? Under what circumstances do teachers change their instructional techniques in response to a significant change in class size?
- What types of secondary resources (e.g., curricular alternatives, supplies and materials, preparation time) aid teachers in their search for instructional strategies that work for them and their students?
- What types of programs or opportunities would induce middle class parents to send their children to urban public schools?
- How do different policies to curb violence in schools influence the behavior of students and, consequently, the learning environment?

In some respects, the research needed to answer these questions is very different from earlier research on the role of school resources in determining children's achievement. The new research agenda focuses on the responses of human resources to incentives provided by institutional rules and to the opportunities and constraints provided by secondary resources. Earlier research treated all school resources as parallel; moreover, it reflected the assumption that resource configurations could be manipulated and "packaged" by officials. This new research agenda pays particular attention to the determinants of resource configurations. In other words, it explores the impact of institutional rules and the quality of secondary resources on the mobility decisions of teachers and families.

While these differences are significant ones, the new research agenda has grown directly out of the earlier research on school effectiveness. Clear evidence from earlier research that schools matter, plus the puzzles created by ambiguous findings on particular resources, led to the perspective developed in this essay. In this respect, the research directions suggested here are a natural successor to the earlier snapshot research.

In time, research on the behavioral responses of teachers, students, and families may enable us to choose public policies with a clear sense of their impact on school effectiveness. However, the research questions are extraordinarily difficult to answer. Consequently, it will be many years before researchers can provide policymakers with reliable predictions concerning the results of particular policy changes in school systems.

Given this situation, it seems important to ask whether there are alternatives to research for taking behavioral responses into account in the decisionmaking process. Lindblom (1959) has argued that the decisionmaking process itself can sometimes solve the problem of developing resource allocation mechanisms that evoke productive, rather than debilitating, behavioral responses. A systematic exposition of this argument is beyond the scope of this paper. However, a brief discussion of teachers' unions and collective bargaining may illustrate the argument.

Effective union leaders know which dimensions of working conditions--for example, class size, preparation periods, protection against violence--are most important to local teachers. They also know what types of resource allocation mechanisms—for example, merit pay—are disliked by their members. The process of collective bargaining reveals these preferences and provides information about their relative importance. When conducted by skilled negotiators in a framework which represents the interests of children and families as well as teachers, collective bargaining can produce resource allocation mechanisms that avoid debilitating behavioral responses (Freeman and Medoff, 1979).

Unions may play a role, not only in articulating preferences, but also in influencing teachers' behavioral responses to new institutional incentives. For example, many districts have introduced early retirement programs in recent years in the hope of inducing older teachers, especially those who are less effective, to retire, thereby reducing the necessity of laying off younger teachers. Some observers have doubted that these programs will succeed because older teachers may react with resentment, feeling that the early retirement choice is an admission that one can no longer function effectively in the classroom. Also, some teachers fear that the existence of an early retirement option could lead to pressure on older teachers to resign. Defensive reactions to this fear could have unexpected and undesirable consequences..

The union can play an important role in facilitating the success of early retirement programs by giving them legitimacy and guaranteeing their integrity. In other words, union support for early retirement programs can give them the status of legitimate benefits, earned through years of service, instead of a dole, distributed to burned out teachers. Moreover, the existence of a well developed grievance procedure can ease fears that the early retirement program would lead to harrassment of older teachers. Thus, the existence of a teachers' union may be important in stimulating constructive responses to policies such as early retirement programs.

The point of the teachers' union example is not to make a blanket rationalization for collective bargaining. It is one of many alternative forms of decision making. Other forms include voting and delegation to professionals. The decisionmaking form that will elicit the most productive behavioral responses will depend on the participants, the issue, and the setting.

The point we would like to emphasize is that choices about decisionmaking forms are extremely important. In our view of the production process for schooling, resources do matter. However, the relationships between the primary inputs, teachers, students, and families, and the outputs, student skills, depend critically on the behavior of the key actors. Their behavior is sensitive to the incentives provided by the school system. Unfortunately, the nature of the responses of these key actors to particular incentives is not well understood. In this view, interest groups such as teachers' unions and parents' associations can play a positive role by providing information about critical behavioral responses, and in some cases, by influencing these responses. Viewed in this perspective, a key policy question is what form of decision making will be most successful in eliciting the critical information about behavioral responses. The effectiveness of public schooling depends to a large extent on our ability to develop and use such decisionmaking processes effectively.

WHAT HAVE WE LEARNED?

The purpose of this essay is to provide an interpretation of school effectiveness research that explains puzzles in the empirical findings and provides a coherent perspective from which to ask new research and policy questions. At this point it may be helpful to recapitulate the basic themes developed in this discussion:

1. There is compelling evidence that schooling makes a difference in determining the cognitive skills of children. Consequently, the search for strategies to make schooling more effective is a worthwhile quest.

2. The primary resources that are consistently related to student achievement are teachers and other students. Other resources affect student achievement primarily through their impact on the attitudes and behaviors of teachers and students.

3. The central school resources—teachers and students—will respond to any changes in the institutional rules, customs, or contract provisions that determine the allocation of resources. Some of these behavioral responses will enhance student achievement; others will diminish achievement. The nature of the responses will depend on the priorities and opportunities of these key actors.

4. Better data and more research will help us to learn more about the relationships between school resources and student achievement in ongoing educational systems. However, quantitative research on school effectiveness, as currently conducted, will not provide reliable information about the effects of changes in resources on student achievement. The reason is that

the methodology does not address the question of how resources are allocated in ongoing systems. Therefore, new approaches need to be developed and applied.

5. A central problem in improving schools is to develop mechanisms for incorporating into the decisionmaking process information about the priorities of the key actors, and consequently about their likely behavioral responses. The quality of public education in the future will be determined not only by the level of resources available, but also by our success in developing policy processes that take into account the behavioral responses of teachers, students, and families.

Footnotes

1. See Hanushek (1979) for a detailed description of the methodology used in school effectiveness research.

2. An alternative to norm-referenced tests is criterion-referenced tests, which are more sensitive to differences in curricula. However, to use such tests to compare curricula or school programs, there must be agreement on the goals of the programs. Murphy and Cohen (1974) document how difficult it is to reach agreement on this issue. The widespread interest in the National Assessment of Educational Progress (NAEP) suggests that it is possible to develop instruments which measure proficiency in a number of skills that are commonly regarded as important. However, it is not coincidental that the design of the data collection in the NAEP prevents analysis of the effectiveness of particular educational programs.

3. Summers and Wolfe (1977) found teaching experience to be negatively related to the achievement of children with low initial achievement. They suggest that this may be due to the fact that the "undampened enthusiasm" of new teachers makes them particularly effective with slow learners, while the skills developed through experience are particularly important in teaching children with above average achievement. This is certainly plausible. However, these results could also be due to a particular type of selection mechanism. Effective experienced teachers may be more likely than ineffective teachers to leave exhausting positions in schools serving large numbers of low achieving children because they face a more attractive opportunity set, both inside and outside the teaching profession. This

selection process could explain the negative relationship between teaching experience and effectiveness in teaching children with low initial achievement. Such selection processes are explained in more detail later in the paper.

4. Henderson, et al.'s interpretation of their peer group results is cited in order to clarify the somewhat complicated nature of these findings. In fact, however, the natural experiment evidence does not provide reliable evidence concerning the effects of a conscious policy of redistributing students. The reason is explained in Section II of this essay.

5. The research surveyed in this essay focuses on resources available at the classroom level. As a result, the role of school principals is not considered. It seems intuitive that principals should also be considered among the primary school resources that affect student achievement.

6. The problem of dysfunctional behavior created by attempts to base compensation on perceived productivity is not unique to public education. Several economists have argued that the strict internal labor market rules that govern resource allocation in many industries are a response to the problems of measuring the productivity of individual workers. See Thurow (1976) and Williamson, Wachter and Harris (1975) for different versions of this argument.

References

Armor, D., et al. *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica: The Rand Corporation, 1976.

Averch, H.A., et al. *How effective is schooling? A critical review and synthesis of research findings*. Santa Monica: The Rand Corporation, 1972.

Berman, P. & McLaughlin, M.W. *Federal programs supporting educational change, vol. VIII. Implementing and sustaining innovations*. Santa Monica: The Rand Corporation, 1978.

Bloom, B.S. Time and learning. *American Psychologist*, 1974, *29*, 682-88.

Chall, J. *Learning to read: The great debate*. New York: McGraw-Hill, 1967.

Freeman, R.B. & Medoff, J.L. The two faces of unionism. *The Public Interest*, 1979, *57*, 69-93.

Glass, G.V. & Smith, M.L. *Meta-analysis of research on the relationship of class-size and achievement*. San Francisco: Far West Laboratory for Educational Research and Development, 1978.

Hanushek, E. Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, 1971, *61*, 280-88.

Hanushek, E. *Education and race*. Lexington, Mass.: D.C. Heath, 1972.

Hanushek, E.A. Conceptual and empirical issues in the estimation of educational production functions. *The Journal of Human Resources*, 1979, summer.

Henderson, V., Mieszkowski, P. & Sauvageau, Y. Peer group effects and educational production functions. Journal of Public Economics, 1978, 10, 97-106.

Jackson, P.W. Life in Classrooms. New York: Holt, Rinehart and Winston, 1968.

Kiesling, H.J. Educator objectives and school production, mimeo, 1979.

Lindblom, C.E. The science of muddling through. Public Administration Review, 1959, 19.

Link, C.R. & Ratledge, E.C. Student perceptions, IQ, and achievement. The Journal of Human Resources, 1979, 14.

Murnane, R.J. The impact of school resources on the learning of inner city children. Cambridge, Mass.: Ballinger Publishing Co., 1975.

Murnane, R.J. & Phillips, B.R. Effective teachers of inner city children: Who they are and what they do, mimeo, 1979.

Murphy, J.R. & Cohen, D.K. Accountability in education--the Michigan experience. The Public Interest, 1974, 36.

Summers, A.A. & Wolfe, B.L. Do schools make a difference. The American Economic Review, 1977, 67.

Thomas, J.A. Resource allocation in classrooms. Final Report to the National Institute of Education, 1977.

Thurow, L. Generating inequality: Mechanisms of distribution in the U.S. economy. New York: Basic Books, 1976.

Van Deusen Lukas, C. Problems in implementing Head Start planned variation models, in Rivlin, A. & Timpane, P.M. (eds.), Planned variation in education. Washington, D.C.: The Brookings Institution, 1975.

Wiley, D.E. & Harnischfeger, A. Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. Educational Researcher, 1974, 3, 7-12.

Williamson, O.E., Wachter, M.L. & Harris, J.E. Understanding the employment relation: The analysis of idiosyncratic exchange. The Bell Journal of Economics, 1975, 6, 250-78.

Winkler, D.R. Educational achievement and school peer group composition. The Journal of Human Resources, 1975, 10, 189-204.

Making Sense of Research on School Effectiveness:

The Primacy of Human Resources in Schooling

by

Richard J. Murnane

Institution for Social and Policy Studies
Yale University

June, 1981

I would like to acknowledge the help and advice of Edward Pauly.

This paper is based on research supported by grant NIE-G-79-0084 from the National Institute of Education.

Recent quantitative studies of school effectiveness have demonstrated that there are significant differences in the amount of learning taking place in different schools and in different classrooms within the same school, even after controlling for the different skills and backgrounds that children bring to school. Until the late 1960s, few policymakers would have found this result interesting. Like most Americans, policymakers believed that schooling mattered and recognized that they and their children learned more in some years of formal schooling than in others. However, in recent years this confidence in the ability of the schools to make a difference has been shaken by the inability of quantitative research to identify consistent relationships between school resources and student achievement. To cite one well known summary of school effectiveness research (Averch et al., 1972):

> Almost every study finds one or two or three school resources that tend to be significantly related to student outcomes. But these studies generally examine a large number of school resources. Along with the two or three resources that are found to be significant many are found to be insignificant. And, when we compare the results of various studies, we find that the same resources do not appear among the lists of significant variables studies have compared (p. 45).
>
> . . . Research has not identified a variant of the existing system that is consistently related to students' educational outcomes (p. 154, italics in original).

Some observers have interpreted the unstable findings on relationships between school resources and student achievement as indicating that schools (and thus new school policies and programs) really do not have the potential to significantly alter children's skill levels. The recent studies showing that there are important differences in the amount of learning taking place in different schools and in different classrooms, even among inner city schools, and even after controlling for the skills children bring to school, are important in rebutting the "schools don't make a difference" interpretation. They provide compelling evidence that whatever the reason for the difficulty in identifying consistent relationships between school resources and student achievement, the reason is not that schools do not make a difference.

The purpose of this paper is to present an alternative explanation for the inconsistent and disappointing results of school effectiveness research. First, research has failed to adequately take account of the fact that the key resources in schooling are human resources—teachers, students, and families. Second, research has not addressed the fact that the learning environment in any classroom is itself the product of decisions made by these key human resources.

The reason this has caused problems for school effectiveness research is that the choices made by teachers, students, and families can substantially alter the quantity and quality of instruction in ways that are very hard to capture with the data on school resources typically used in school effectiveness studies. This theme is illustrated by discussing research on two school resources that are of particular interest to policymakers: class size and teacher experience.

Class Size

The impact of class size on student achievement has been one of the most confusing issues in educational research. Despite the efforts of many researchers and the utilization of increasingly large and detailed data sets, no consensus has been achieved on the role that class size plays in determining student achievement. A recent synthesis of past research by Glass and Smith (1978) found that average class size was negatively related to student achievement in classes with fewer than 20 students. However, the Glass and Smith study also reported no significant difference in student achievement between classes with 25 and 35 students.

This finding runs counter to the intuition of many educators, who believe that they can do a better job in helping children to learn if they do not need to spread their efforts over a large number of children. If this is true, why doesn't the evidence on class size reflect this?

One reason it has been difficult to pin down the effects of class size is confusion concerning the definition of class size. Should class size refer to the number of children in the class on any given day? Or should it refer to the number of different children a teacher must serve during a school year? In schools in which there is no turnover among students, there is no difference between these two definitions. However, in schools serving highly mobile student populations, the number of children in membership in a class on any given day may be much smaller than the total number of students the teacher serves during the school year. In such schools, teachers are continually faced with the problem of integrating new children into the class. This task imposes large demands on teacher time and reduces the time available for instruction of the rest of the class. Consequently, in classes in which there is a significant significant amount of student turnover, the number of students in the class at any one time may not reflect the demands on the teacher's time, and consequently may not reflect the amount of instruction received by the students who do stay in the class for the entire year.

Thus, in effect there are two alternative dimensions of class size, the average number of students in membership on any day, and the total number of different students that belong to the class during the school year. Holding average class size constant, the total number of students will be larger, the great the amount of student turnover. Each of these dimensions of class size may have an impact on student achievement.

In a study based on information on a sample of 800 inner city elementary school children and their teachers, Murnane (1975, 1981) examined the impact of these two dimensions of class size on student achievement. The results showed that the average number of students in a class was not significantly related to student achievement. (This may have been due to the limited variation in average class size in the sample.) However, the total number of students who passed through the class during the school year was negatively related to student achievement. In other words, the greater the amount of student turnover in a class, the lower the achievement of the children who did stay in that class for the entire year. This supports the hypothesis that the need to continually integrate new children into a class during the school year reduces the amount of time available to instruct the stable student population.

The key lesson from this study is that average class size may not reflect accurately the demands on a teacher's time and the amount of instruction provided to children in schools serving transient student populations.

Teacher Experience

The relationship between teacher experience and teacher effectiveness is another heavily researched issue for which the evidence remains inconclusive. Some studies report positive relationships between teaching experience and teaching performance, as measured by student achievement gains (Hanushek, 1972; Murnane, 1975; Kean et al., 1979). Other studies report no significant relationship (Hanushek, 1971; Henderson et al., 1978; Link and Ratledge, 1979). This section explains that the puzzling nature of the evidence is due at least in part to the research methodology used to study the experience-performance relationship.

The hypothesis that teachers become more effective as they gain experience rests on the view that teaching is a complex process requiring a varied set of skills, many of which can only be learned on the job. In other words, teachers learn to teach by teaching and as a result they become more effective as they acquire experience. The most straightforward way to investigate the impact of learning by doing on teaching performance is to examine the effectiveness of individual teachers over time. To date, this strategy has not been used.

Instead, the role of learning by doing has been investigated by estimating the relationship between experience and performance for a sample of teachers at one point in time. It has been implicitly assumed that, after taking into account observable differences among teachers such as the quality of the college they attended, the only remaining reason that teachers differ in effectiveness is as a result of differences in experience. However, there are good reasons why there may be important unobserved differences in the effectiveness of teachers with different levels of experience that are not the result of learning learning by doing. These differences are the result of vintage and self-selection effects.

Vintage effects are differences in the average abilities of teachers hired by school districts at different points in time. The most compelling explanation for the existence of vintage effects is that dramatic changes in labor market conditions for teachers over the last twenty five years have affected the quality of new entrants to the teaching profession. In the late

1950's and early 1960's a rapid increase in student enrollments created an acute shortage of teachers in the United States. Many school districts, particularly urban districts, found it difficult to find qualified applicants to fill vacant positions. By 1970 this situation had changed significantly. Due to the combination of a decrease in the demand for teachers precipitated by declining enrollments and an increase in the supply of teachers (a delayed response to the earlier shortage), there has been a surplus of teachers in most subject areas during the 1970's. As a result, school districts have been able to be very selective in choosing among the large number of applicants for teaching positions. Assuming that district personnel officers are able to identify applicants with the greatest potential, the average quality of new teachers should be higher in periods of excess supply than in periods of excess demand.

Self-selection is another reason that teachers with different amounts of experience at a given point in time may differ in effectiveness. The self-selection hypothesis states that the effectiveness of teachers who choose to remain in the profession (or in a particular school district) may differ systematically from the effectiveness of teachers who choose to leave. A variety of mechanisms could create these selection effects. For example, the more effective experienced teachers may leave the classroom to become administrators. Similarly, it may be only the most able teachers who survive the difficult first years of teaching. It could also be that effective teachers may be the most likely to leave teaching to pursue occupations in which high skill levels are rewarded with especially high salaries. Self-selection and vintage effects both influence the relationship between years of teaching experience and teaching performance in a sample of teachers observed at one point in time and confound attempts to assess the impact of learning by doing on teaching performance.

In an attempt to improve the methodology used to study the impact of learning by doing on teaching performance, Murnane and Phillips (1981) investigated how sensitive estimates of the impact of learning by doing on teaching performance were to vintage effects. They found that teaching experience was not significantly related to teaching performance when vintage effects were not taken into account. However, when vintage effects were taken into account, teaching experience was positively related to teaching effectiveness. In other words, the impact of learning by doing on teaching performance could only be observed when vintage effects—differences in the abilities of teachers resulting from changing labor market conditions—were taken into account.

Conclusion

The point of this article is to explain one important reason why school effectiveness research has not identified stable relationships between school resources and student achievement: namely, that the school resources observable in a classroom at one point in time may not accurately reflect the quantity and quality of instruction received by the students in that class. For example, the number of students in the class on any given day may not reflect the demands on the teacher's time, if there has been significant turnover of students during the school year. The number of years of experience a teacher has had may not reflect the quality of the teacher's performance, if the influence of learning by doing is countered by vintage and selection effects.

The lesson for researchers and managers is that the key resources in the schooling process, families, students, and teachers, are continually making decisions that determine the quality of the learning environment in schools and classes. For example, family decisions about whether to move during the school year influence the amount of turnover among the students of a school and consequently influence the amount of time available for instruction. The decisions of talented college graduates about whether to enter the teaching profession and how long to stay in the profession influence the quality of instruction children receive.

While this lesson may seem obvious, it has often been forgotten by researchers and school managers alike. Too often researchers fail to ask what the data they have collected on class size and teacher experience really reveal about the quantity and quality of instruction children receive, in a world in which transient families and changing labor market conditions for teachers are important facts.

School managers, who are under great pressure to use resources efficiently and to provide ready answers to a demanding public, also sometimes forget that the key school resources are people who respond to any and every policy change. Great care must be taken to assure that the responses of the key human resources to a policy change will in fact enhance the quality of instruction.

In summary, to make sense of the results of school effectiveness research and to use these results effectively, both researchers and policymakers need to focus on the central role that human resources play in schooling.

References

Averch, H.A. et al. How Effective is Schooling? A Critical Review and Synthesis of Research Findings. Santa Monica: The Rand Corporation, 1972.

Glass, G.V. and M.L. Smith. Meta-analysis of Research on the Relationship of Class Size and Achievement. San Francisco: Far West Laboratory for Educational Research and Development, 1978.

Hanushek, E. Education and Race. Lexington: D.C. Heath, 1972.

Hanushek, E. "Teacher characteristics and gains in student achievement: Estimation using micro data." American Economic Review, 1971, 61, 280-88.

Henderson, V. et al. "Peer group effects and educational production functions." Journal of Public Economics, 1978, 10, 97-106.

Kean, M.H. et al. "What works in reading?" Philadelphia: Office of Research and Evaluation, Philadelphia School District, 1979.

Link, C.R. and E.C. Ratledge. "Student perceptions, IQ and achievement." Journal of Human Resources, 1979, 14, 98-111.

Murnane, R.J. The Impact of School Resources on the Learning of Inner City Children. Cambridge: Ballinger, 1975.

Murnane, R.J. "Student turnover: A missing piece of the class size puzzle." Mimeo, 1981.

Murnane R.J. and B. Phillips. "What do effective teachers of inner city children have in common." Social Science Research, 1981, 10.

Input-Output Research in Education:
Accomplishments, Limitations, and Lessons

by

Richard J. Murnane

Institution for Social and Policy Studies
Yale University

March, 1982

This paper is based on research supported by grant, NIE-G-79-0084 from the National Institute of Education.

Input-output research has made valuable contributions to our understanding of schools; however, it does not provide reliable evidence concerning how school resources should be allocated.

Input-Output Research in Education:
Accomplishments, Limitations, and Lessons

Richard J. Murnane

I. INTRODUCTION

Over the last 16 years, quantitative research on school effectiveness--called educational production function studies by economists, input-output studies by sociologists, and research on the cost-quality issue by lawyers--has played a significant role in public policy debates concerning a range of educational issues. The results of school effectiveness studies have been introduced into court cases dealing with the way schools are financed, into legislative debates concerning compensatory education, and into executive branch deliberations concerning school busing.

The prominent role that this research has played in policy discussions has led many observers to ask: How good is the research? What does it really have to tell us? Does it provide reliable evidence concerning how scarce resources should be allocated? The purpose of this paper is to address these questions. Section II describes the accomplishments of quantitative research on school effectiveness, including advances in our understanding of what the critical resources in schooling are. Section III explains the limitations of this research--in particular, why it does not provide a basis for determining how school resources should be allocated. This argument is illustrated with two extended examples, describing research on the effects of teaching experience and class size on school effectiveness. Section IV discusses lessons that follow from the arguments presented here. One of the lessons--the importance of talking to school people about alternative explanations of empirical findings--has the potential for increasing dialogue between researchers and school officials and for throwing light on the actual school policies that influence student achievement.

II. ACCOMPLISHMENTS

In the 16 years since the publication of Equality of Educational Opportunity, the first large-scale quantitative study of school effectiveness, important advances have been made in the methodology used to study the relationship of school resources to student achievement. These advances include the use of longitudinal data measuring student gains in skills from one school year to the next, and the use of more accurate measures of the school resources actually available to individual students. As a result of these methodological advances, quantitative studies of school effectiveness, which I will refer to as input-output studies, can now provide quite detailed descriptions of relationships that exist between school resources and student achievement over the course of a school year.

What have these descriptions revealed? The most important finding is that there are significant differences in the amount of learning taking place in different schools and in different classrooms within the same school, even among inner city schools, and even after taking into account the skills and backgrounds that children bring to school (Armor et al., 1976; Hanushek, 1971; Murnane, 1975). Until the late 1960s, few policymakers would have found this result interesting. Like most Americans, policymakers believed that schooling mattered and recognized that they and their children learned more in some years of formal schooling than in other years. However, in recent years frustration resulting from the difficulties in finding "policy levers" to improve schools, and from disappointing evaluations of policy innovations that were tried, has shaken this confidence. The results indicating clear differences among schools and classrooms affirms that it is worthwhile devoting attention to the fact that some schools provide better education than other schools do, despite our limited success in tracing this fact to policies responsible for it.

The second important set of positive findings from the quantitative research on school effectiveness is the critical importance of the schools' human resources, teachers and students. (Due to data limitations, no large-scale input-output studies to date have investigated the role of the school principal. However, school effectiveness studies using other kinds of research designs (Edmonds, 1979) indicate that school principals play a central role in determining school effectiveness.)

Virtually all studies find that some attributes of teachers are significantly related to student achievement. The attributes most commonly related to student achievement are those providing information about teachers' intellectual skills, such as scores on tests of verbal ability, or the quality of the college a teacher attended (Hanushek, 1981; 1979; Murnane, 1981; Summers and Wolfe, 1977, Winkler, 1975).

Many studies also report the importance of student body composition. While the results vary somewhat across studies, there is support for the hypothesis that elementary school children with low initial skill levels who attend schools in which their classmates' average achievement level is relatively high make more progress than such children who attend schools in which their classmates' average achievement level is relatively low. There is similar evidence regarding socio-economic status (Henderson et al., 1978; Summers and Wolfe, 1977; Winkler, 1975).

To appreciate the contribution of recent input-output research to our understanding of schools, it is helpful to place this research in a historical perspective. Quantitative research on school effectiveness began with a broadly specified model that was agnostic on the roles played by particular school resources. In the model, a large number of resources were treated in parallel fashion, including physical facilities such as the number of library books and the size and age of the school, as well as human resources. We now understand that the primary resources are teachers and students. If other resources matter, it is through their impact on the behaviors of teachers and students.

III. LIMITATIONS

While there has been growing agreement among researchers and policymakers concerning which school resources are most important (Hanushek, 1979; Murnane, 1981), no such agreement exists concerning the policy implications of these findings. Some economists have suggested that the results indicate how school resources should be allocated, and have formalized this suggestion by incorporating the results of input-output studies into optimizing algorithms designed to show how school districts should allocate resources to reach particular goals (Cohn, 1978; Boardman, 1980). Other economists, while not stating what school officials should do with dollars, argue that schools must be using resources inefficiently since the pattern of payments of resources does not reflect the relative values of the regression coefficients in input-output studies (Hanushek, 1981).

Many school officials view these interpretation of the evidence with skepticism. They feel intuitively that the quantitative research does not capture all of the information needed for good policymaking. Moreover, while sometimes conceding that resources could be used more effectively, most school officials believe that dollars as currently used are related to school quality. However, since most school officials are not trained researchers, they are usually unable to express the sources of their skepticism in the language of models and statistics. At the same time, few researchers have tried to elicit from school officials ideas about aspects of schooling that should be reflected in the design of quantitative research. Thus, as a result of differences in training and inclinations, little dialogue takes place between quantitative researchers and school officials.

In this section I explain the methodological limitations of input-output research in education that I believe underlie the skepticism of school officials. These limitations make it inappropriate to base decisions about how school resources should be allocated on the results of this research.

Input-output research essentially provides snapshots—albeit increasingly fine-grained snapshots—of relationships between school resources and student achievement. To serve as an adequate basis for resource allocation decisions, these snapshots must fulfill two conditions:

first, they must provide accurate information about the (unobserved) moving picture that constitutes the school process;

second, they must accurately predict how the moving picture would be altered, and ultimately how student achievement would be affected, by changes in allocations of school resources.

It is my contention that input-output research in education does not fulfill these conditions, and thus cannot tell us how to allocate school resources. To clarify the reasons, it is useful to contrast input-output research in education with similar research in agriculture.

In an oft cited paper, Heady (1957) reported on his investigations of the optimal combinations of fertilizers—nitrogen and phosphate—that farmers should use in growing corn. Heady began his research by conducting controlled experiments to determine the effects of different input levels of the fertilizers on corn yields. He then analyzed his experimental data with multivariate methods to estimate the input-output relationships between fertilizer combinations and corn yield. Finally, Heady incorporated information on the prices of the two inputs with the results of the input-output study to calculate how farmers should combine inputs of nitrogen and phosphate to maximize the yield from a given level of expenditures on fertilizer.

The critical characteristics of the agricultural research example that allowed the research results to serve as a reasonably reliable policy guide for farmers are the following:

1. Heady did controlled experiments to assure that his results did in fact reflect the causal influences of the fertilizers on corn yields.
2. The inputs were well-defined homogeneous commodities available in well-operating markets at clearly identified prices.
3. The input mix could be changed by simply altering the amounts of nitrogen and phosphate that were purchased and spread on the corn fields.

Contrast these characteristics with the situation facing the researcher investigating input-output relations in education:

1. The researcher must rely on data from natural experiments—that is, from the variation in resource combinations that naturally occurs in schools. This is a significant limitation in assessing causation because not only are the critical school resources, teachers and students, not assigned to each other by a random assignment process, but in fact a variety of formal and informal assignment procedures exist that tend to group

students with particular characteristics. As a result it is extremely difficult to disentangle input-output relationships from the influence of assignment procedures.

2. In input-output research in education, the critical resources, teachers and students, are characterized by their attributes (for example, the race, sex, experience, and education of teachers and the socio-economic background and prior skill levels of students). However, it is not possible for schools to acquire individual attributes, only bundles of correlated attributes. Moreover, the bundles cannot be purchased in well-defined markets at clearly defined prices. Instead, the acquisition process consists of defining personnel policies and student attendance policies to which teachers and students respond.

3. Changing the resource allocation mix—that is, the combinations of teachers and students that work together--requires changing personnel policies and/or student attendance policies. These changes will elicit responses from teachers and students that are often unpredicted. (Think, for example, about the responses elicited by some busing programs (Rossell and Hawley, 1981) and teacher layoff programs (Johnson, 1980).) Input-output research, as currently conducted, does not examine the nature of these responses because it takes as given the combination of teachers and students that are present in different classrooms and schools. Consequently, the resource allocation process is beyond the observation and control of the analyst.

School officials, as a result of their experiences and frustrations in trying to create effective school programs, are aware at least implicitly of some of these factors that so powerfully differentiate the education case from the agriculture case. In particular, they know how difficult it is to change personnel and student attendance policies and they know that policy changes often elicit responses very different from the expected responses. It is these experiences that produce the skepticism of many school officials concerning the policy implications of input-output studies.

The next two subsections provide examples of research puzzles that reflect the difficulty of capturing critical elements of the moving picture of schooling with input-output research. In both examples, altering the research design in ways that reflected the intuitions and observations of school officials diminished the puzzling nature of the results and increased the extent to which the research provided insights about the ways that the actions of teachers and students affect student achievement.

A. Teacher Experience. The relationship between teaching experience and teaching effectiveness is one of the most heavily researched questions in education. However, despite many studies, the evidence remains inconclusive. Some studies report positive relationships between teaching experience and teaching performance, as measured by student achievement gains (Hanushek, 1972; Kean et al., 1979; Kiesling, 1981). One study reports a negative relationship between teacher experience and effectiveness in teaching students with certain characteristics (Summers and Wolfe, 1977). Many other studies report no significant relationship between teaching experience and performance (Armor et al., 1976; Hanushek, 1981; Henderson et al., 1978; Link and Ratledge, 1979).

Many school officials find these results puzzling. From their observations of individual teachers over time, they conclude that most teachers learn a great deal about how to teach in their first years on the job and that this learning by doing results in improved teaching.

In fact, the puzzling and conflicting nature of the research evidence is due at least in part to the research methodology used to study the experience-performance relationship. To date, input-output studies have not investigated the impact of learning by doing on teaching performance by the most reliable method, namely, by examining how the performance of individual teachers changes as the teachers acquire experience. Instead, the studies have investigated the relationship between experience and performance for a sample of teachers at one point in time. It has been implicitly assumed that, after taking into account observable differences among teachers (such as the quality of the college they attended), the only remaining reason that teachers differ in effectiveness is as a result of differences in learning by doing. However, there may be important unobserved differences in the effectiveness of teachers hired in different years that are not the result of learning by doing. These differences are the result of vintage and self-selection effects.

Vintage effects are differences in the average abilities of teachers hired by school districts at different times—for example, in different years. The most compelling explanation for the existence of vintage effects is that dramatic changes in labor market conditions for teachers over the last 25 years have affected the quality of new entrants to the teaching profession. In the last 1950s and early 1960s, a rapid increase in student enrollments created an acute shortage of teachers in the United States. Many school districts, particularly urban districts, found it difficult to find qualified applicants to fill vacant positions. By 1970 this situation had changed significantly. Due to the combination of a decrease in the demand for teachers precipitated by declining enrollments and an increase in the supply of teachers (a delayed response to the earlier shortage), since 1970 there has been a surplus of teachers in most subject areas. As a result, school districts have been able to be very selective in choosing among a the large number of applicants for teaching positions. Assuming that district personnel officers are able to identify applicants with the greatest potential, the average quality of new teachers should be higher in periods of excess supply than in periods of excess demand. Unless the differences in the average abilities of teachers hired at different times are captured by variables describing teacher backgrounds (which is very difficult to do), research based on a cross-section of teachers will not produce reliable estimates of the influence of teaching experience on teaching performance.

Self-selection is another explanation for the puzzling results of cross-sectional research on the relationship between teaching experience and teaching performance. The self-selection hypothesis states that the effectiveness of teachers who choose to remain in the profession (or in a particular school district) may differ systematically from the effectiveness of teachers who choose to leave. For example, it may be only the most able teachers who survive the difficult first years of teaching. Conversely, effective experienced teachers may leave the classroom to become administrators. Or effective teachers may be the most likely to leave teaching to pursue occupations in which high skill levels are rewarded with especially high salaries. Self-selection, like vintage effects, confounds attempts to use cross-sectional data on teachers to assess the impact of learning by doing on teaching performance.

In an attempt to improve the methodology used to estimate the impact of experience on teaching performance, Murnane and Phillips (1981) constructed an explicit measure of vintage effects for a sample of teachers in one large urban school district and investigated how sensitive the estimates of the experience-performance relationship were to the inclusion of this measure. The measure of vintage effects consisted of the change in total student enrollments in the school district between the year in which the teacher first taught in the district and the previous year. The logic underlying the use of this measure is that in years of rapidly growing student enrollments, this district and neighboring districts competing for teachers in the same labor market hired large numbers of teachers at a time in which the supply of teachers was relatively limited. As a result, they could not be selective in choosing among applicants. When enrollments were declining, personnel officers could be more selective and average teacher quality would rise.

The variable included in the model to indicate learning by doing was the natural logarithm of each teacher's total number of years of teaching experience. This specification reflected the assumption that teachers continue to learn as they gain experience, but that the greatest gains from additional experience occur in the first years of teaching.

The empirical results indicated that teaching experience was not significantly related to teaching performance when vintage effects were not taken into account. However, when vintage effects were taken into account, teaching experience was positively related to teaching experience. The size of the relevant coefficient implied that children taught by a teacher with five years of experience made three to four more months of progress in acquiring reading skills during a school year than did children taught by a first year teacher.

The key lesson from this example is that the career decisions of teachers, made in the context of changing labor market conditions for teachers, make it extremely difficult to capture with cross-sectional data an important element of the moving picture of schooling—namely, how the performance of teachers changes as they gain experience.

B. Class Size. The impact of class size on student achievement has been another confusing issue in educational research. Despite the efforts of many researchers and the utilization of increasingly large and detailed data sets, no consensus has been achieved on the role that class size plays in determining student achievement. Many educators are skeptical about the inconclusive research findings because their experiences suggest that most teachers do a better job in helping children to learn when they do not need to spread their efforts over a large number of children.

As was the case with the experience-performance relationship, one reason for the ambiguous research results involves the difficulty of capturing critical aspects of the moving picture of schooling with essentially snapshot research. In the case of class size research, this difficulty is reflected in confusion about the appropriate definition of the concept. Should class size refer to the number of children in the class on any given day? Or should it refer to the number of different children a teacher must serve during a school year? In schools in which there is no turnover among students, there is no difference between these two definitions. However, as the principal of one urban elementary school explained to me, in schools serving highly mobile student populations, the number of children in membership in a class on any given day may be much smaller than the total number of students the teacher serves during the school year. In such schools, teachers are continually faced with the problem of integrating new children into the class. This task imposes large demands on teacher time and reduces the time available for instruction of the rest of the class. Consequently, in classes in which there is a significant amount of student turnover, the number of students in the class at any one time may not reflect the demands on the teacher's time, and consequently may not reflect the amount of instruction received by the students who do stay in the class for the entire year.

Thus, in effect there are two alternative dimensions of class size, the average number of students in membership on any day, and the total number of different students that belong to the class during the school year, Holding average class size constant, the total number of students will be larger, the greater the amount of student turnover. Each of these dimensions of class size may have an impact on student achievement.

In a study based on information on a sample of 800 inner city elementary school children and their teachers, Murnane (1975) examined the impact on student achievement of these two dimensions of class size. Average class size was measured by the average of the number of students enrolled in a class on October 15 and April 15. The total number of students served was measured by the total number of names that appeared on the class register for that school year.

The results indicated that average class size was not significantly related to student achievement. (This may have been due to the limited variation in average class size in the sample.) However, the total number of students who passed through the class, which varied from one more than average class size to 22 more than average class size, was negatively related to student achievement.

The primary lesson from this example is similar to the most important lesson from the previous example—namely, that the active behaviors of the human resources make it difficult to capture critical elements of the schooling process with cross-sectional data. In this case, it was the mobility of students that made average class size a poor measure of the demands on teachers' time and confounded attempts to investigate whether the level of these demands affected student achievement.

IV. LESSONS

One lesson to be drawn from the two input-output studies described in the previous section is that the intuitions of experienced school officials should be viewed as significant sources of illumination to which researchers should pay attention in designing input-output research. Often school officials' perceptions about what affects student achievement in their schools can be incorporated into input-output research, improving the ability of the research to capture critical dimensions of the ways schooling affects students' achievement. For example, in the experience and class size research described above, interviews with school officials produced the ideas for studying vintage effects and student turnover.

A second lesson is that it is important to consider a broad set of possible interpretations of coefficients in input-output studies. In other words, researchers should think carefully about alternative explanations of significant coefficients, and view as highly plausible and also potentially interesting, explanations other than the usual explanations that the relevant right-hand side variable had a causal influence on student achievement.

The reason why alternative explanations are highly plausible is that the resource combinations present in particular classrooms are determined by the large number of personnel and student assignment policies, explicit and implicit, present in all school districts, and by the reactions of students and teachers to these policies. Policies and the reactions of the human resources to policies influence student achievement in many different and often unpredictable ways. These influences often result in significant coefficients in input-output studies, but frequently the influences do not reflect causal relationships between the observed explanatory variables and student achievement.

Careful researchers do try to be sensitive to these alternative explanations and many published articles describing the results of input-output studies contain footnotes mentioning alternative explanations of particular significant coefficients. However, the typical relegation of these alternative explanations to the footnotes suggests that most researchers view them as problems obstructing the estimation of the central input-output relationship rather than as interesting and potentially important phenomena in their own right.

This perspective is also indicated by the language used in the econometric literature to describe such alternative explanations--selectivity bias, omitted variables, nonrandom samples. These terms connote problems that obstruct researchers' attempts to examine a particular well-defined concept, such as an input-output relationship. In cases like the fertilizer-corn yield example (in which the input-output relationship is clearly defined, is independent of resource allocations mechanisms, and is the one critical puzzle), this perspective is reasonable. It keeps the researcher focused on the one central question. In the fertilizer-corn yield case, focusing exclusively on the input-output relationship was possible and appropriate. Moreover, it was sufficient to answer the policy question of what to do with scarce resources because no signficant puzzles or process difficulties were involved in buying and allocating the relevant inputs.

The education case is quite different, however. The critical resources, teachers and students, cannot simply be purchased and placed in classrooms. Instead, changes in human resource combinations can only be brought about by changing personnel or student attendance policies. Often the intended changes in resource combinations do not occur because teachers and/or students respond to the policy changes in unpredicted ways. In other words, teachers' and students' reactions to the many formal and informal rules and policies that characterize schools are aspects of the motion picture of schooling that are not well understood and are critical in determining how schools influence students. Pursuing alternative explanations of puzzling results of input-output studies is one potentially fruitful way of learning more about these poorly understood aspects.

To illustrate this suggestion, consider two examples of findings from recent input-output studies that are somewhat puzzling and have many alternative explanations.

Murnane (1975) found that black teachers in an urban school system were more effective in teaching reading skills to black primary school children than white teachers in the school system were. Among the conventional explanations consistent with a causal interpretation are that black teachers may understand the needs and motivational patterns of black students better than white teachers do and that black teachers may provide a role model that inspires black students to greater effort. There are also alternative possible explanations of a somewhat different nature. Most of the black teachers in the sample were educated at one of a small number of black colleges in the South. As a result of personal connections with placement officers in these colleges, the school district personnel director might have been able to recruit the most able graduates of these schools. Another possibility is that as a result of differential alternative employment possibilities, talented black teachers may have been less likely to leave the school system after a few years of teaching than talented white teachers were. The latter two explanations concern idiosyncratic aspects of the operation of labor markets that are important to understand in designing policies to recruit and retain talented teachers. Interviewing the personnel director and the relevant black teachers might have provided information about the relevance of these alternative explanations.

A second example concerns Summers and Wolfe's (1977) counterintuitive finding that for students who began the school year with average or below average achievement levels, the number of disruptive incidents that occurred in the school during the school year was positively associated with student achievement growth. To the authors' credit, the disruptions variable was not deleted from the analysis and one possible explanation of the finding was discussed in the text of the article—namely, that the results might stem from the aggregated nature of the disruption data. However, might there not be alternative explanations that reflect real things going on in schools, rather than simply artifacts of data limitations? Possibilities include that in such schools only atypical students take achievement tests (a necessary condition for inclusion in the sample) or that administrators have found ways to attract particularly effective teachers to troubled schools.

Undoubtedly there are a variety of other possible explanations and the point of this discussion is not to defend any particular candidate.

Instead, the point is to argue that exploring diligently a variety of possible explanations for puzzling results may uncover either intended or unintended aspects of the schooling process that are important in determining students' achievement.

This type of research requires methods such as interviewing school officials that researchers specializing in quantitative analysis generally are not comfortable with. However, conducting interviews with the explicit objective of unraveling puzzles posed by the results of quantitative research may greatly enhance the contributions of quantitative research in describing important elements of the schooling process, and ultimately in providing information that can lead to better schools.

REFERENCES

Armor, D., and others. Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools. Santa Monica: The Rand Corporation, 1976.

Boardman, A.E. "Policy Models for the Management of Student Achievement and Other Educational Outputs." Management Science, 1978, 8, 99-122.

Cohn, E. "Combining Efficiency and Equity: Optimization of Resource Allocation in State School Systems." In W.W. McMahon and T.G. Geske (Eds.), Toward Efficiency and Equity in Educational Finance. Boston: Allyn and Bacon, 1980.

Coleman, J.S., and others. Equality of Educational Opportunity. Washington, D.C.: Office of Education, National Center for Education Statistics, 1966.

Edmonds, R.R. "Some Schools Work and More Can." Social Policy, 1979, 9 (5), 28-32.

Hanushek, E.A. "Throwing Money at Schools." Journal of Policy Analysis and Management, 1981, 1 (1), 19-41.

________. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions." Journal of Human Resources, 1979, 14, 351-388.

________. Education and Race. Lexington: D.C. Heath, 1972.

________. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." American Economic Review, 1971, 61, 280-288.

Heady, E.O. "An Econometric Investigation of the Technology of

Agricultural Production Functions." Econometrica, 1957, 25, 249-268.

Henderson, V., Mieszkowski, P., and Sauvageau, Y. "Peer Group Effects and Educational Production Functions." Journal of Public Economics, 1978, 10, 97-106.

Johnson, S.M. "Performance-based Staff Layoffs in the Public Schools: Implementation and Outcomes." Harvard Educational Review, 1980, 50, 214-233.

Kean, M.N., Summers, A.A., Raivetz, J.J., and Farber, I.J. "What Works in Reading?" and "What Works in Reading--Technical Supplement." Office of Research and Evaluation, School District of Philadelphia, 1979.

Kiesling, J. "The Production of Reading." Indiana University, mimeo, 1981.

Link, C.R. and Ratledge, E.C. "Student Perceptions, IQ, and Achievement." The Journal of Human Resources, 1979, 14 (1), 98-111.

Murnane, R.J. "Interpreting the Evidence on School Effectiveness." Teachers College Record, 1981, 83 (1), 19-35.

________. The Impact of School Resources on the Learning of Inner City Children. Cambridge: Ballinger, 1975.

Murnane, R.J., and Phillips, B. "Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance." Economics of Education Review, 1981, 1 (4), forthcoming.

Rossell, C., and Hawley, W. "Understanding White Flight and Doing

Something About It." In W. Hawley, Effective School Desegregation. Beverly Hills: Sage Publications, 1981.

Summers, A.A., and Wolfe, B.L. "Do Schools Make a Difference?" American Economic Review, 1977, 67 (4), 639-652.

Winkler, D.R. "Educational Achievement and School Peer Group Composition." The Journal of Human Resources, 1975, 10, 189-204.

Input-Output Relations and Innovation in Education

by

Richard J. Murnane[*]

and

Richard R. Nelson[**]

Institution for Social and Policy Studies
Yale University

March, 1982

---

[*] Research was supported by Grant NIE-G-79-0084 from the National Institute of Education.

[**] Research was supported by grants from the Sloan Foundation and the National Science Foundation.

ABSTRACT

This article is concerned with the applicability of the conventional economic theory of production to activities such as education. Part I describes the conditions under which the conventional theory provides a useful theoretical framework for understanding input-output relations in a sector. Parts II and III explain why existing knowledge about input-output relations in education and changes in these relations over time is not compatible with the assumptions underlying conventional theory. Part IV provides suggestions toward the development of a theory of production that is applicable to activities such as education.

I. INTRODUCTION

We take it as tautological that the output of any organization can be exactly explained by the inputs and techniques it is using, if these are appropriately defined. It follows that differences in outputs among organizations, and over time, also can be explained in terms of differences in inputs and techniques.

Recognition of logically necessary relations, often in the form of accounting identities, is an important aspect of human understanding. However, tautologies, like MV equals PT, or Y equals C plus I, take on analytic life, and become part of a theory, only as various non-tautological assumptions are breathed into them. Thus the assumption that V is a constant, or that C equals c times Y, transforms the identity into a theory that asserts relationships not innate in the tautology and which, therefore, may be basically right or basically wrong. Similarly, the assumptions implicit in the standard economic theory of production breathe analytic life, and the potential for illumination or obfuscation, into the output, input, technique, tautology.

The standard economic theory of production was developed and employed initially in the context of the broader theory of the profit maximizing firm. When it is used outside that context (e.g., in the study of input-output relations and innovations in education) the general assumptions of the mother context tend to be carried over to the offspring. While seldom spelled out explicitly, three important (non-tautological) assumptions are employed in most economic analyses of production.

When it is used outside that context (e.g., in the study of input-output relations and innovations in education) the general assumptions of the mother context tend to be carried over to the offspring. While seldom spelled out explicitly, three important (non-tautological) assumptions are employed in most economic analyses of production.

1. The inputs and techniques employed by a firm are rationally chosen from a set of well-defined possibilities, so as best to achieve the organization's objectives. While more complex versions of the theory may admit some factors not under the decision maker's control--e.g., the amount of rainfall--the basic flavor is that the values of the right-hand side variables can be explained in terms of rational choice by a decision maker.
2. The requisite inputs and techniques are generally available. While special site advantages, instances of unusual skill, or even proprietary knowledge, may be recognized in some cases, these idiosyncratic elements are not stressed.
3. While it may be recognized that considerable trial and feedback learning may be required before a good technique-input combination tailored to the situation is found, it is presumed that most observations of inputs and outputs are of situations where this exploration has been essentially completed.

All three of these assumptions show through in specially strong form in research which aims to explain differences in outputs and inputs among firms in terms of different points along a common production function. Varying market conditions are presumed to lie behind the variation of the points; the points on the production function are assumed to have been chosen so as to maximize profit. Differences in access to particular inputs generally are ignored in such analysis. Similarly, the presumed existence of a common production function makes sense only if all techniques are available and known to all firms. Finally, the presumption that firms are on the production function means that experimentation to find it is essentially over.

Behind the scenes (usually) is an implicit theory of the nature of techniques. They are well articulated, like the recipes in a cookbook. The ingredients referred to in the recipe can be bought at most stores. Such operations as mixing or baking can be accomplished with generally available spoons and stoves. Thus "recipes" indeed can be chosen, by anyone who has the relevant cookbooks, and can be performed by anyone with adequate cooking skills and access to the relevant stores.

In the standard theory of production it is assumed that everyone has access to all cookbooks, and "knows" what is in them. In some applications of the theory of production, as in the studies of education effectiveness we shall review in this paper, it is not assumed that all techniques are known, or available, to all organizations. In such studies the techniques employed themselves may be entered as arguments of the production function. As we shall elaborate later, these studies usually carry the connotation of easy "technology transfer." Absent legal constraints, and perhaps with a bit of teaching and learning, any organization could use any technique with roughly comparable results. Thus the basic presumptions about the nature of "techniques" in education are consistent with that in the theory of the firm. Technological knowledge is assumed to be well-articulated and to define, quite precisely, both what can be done and how to do it. The "how to do it" part involves acts that any reasonably skilled person can perform with inputs that are generally available.

This implicit theory about techniques often is complemented by a set of assumptions about new technique generation.

4. The advance of technological knowledge (the creation of new techniques) is presumed to occur through activities involving specially trained people who do their work at some distance from actual production. This separate activity might as well be called _research_ and _development_. Since the new techniques are perfectly articulated, there is no particular problem of transference of knowledge from the R&D laboratory to actual use.

These assumptions about technological advance are explicit or implicit in many models of productivity growth in manufacturing industries. As we shall see, many scholars of education consider educational innovations as coming about largely through educational R&D.

We have stated the assumptions that we think like behind the standard economic theory of production, technology, and technological advance quite starkly. While some may quibble about the details, we maintain that these assumptions, or notions quite close to them, are needed if the concept of a production function is to have much meaning, if it is appropriate to think of all organizations as capable of doing the same things, and if the source of enhanced production capabilities is meaningfully to be sought in research and development. Further, as we shall argue later on in this paper, these notions are implicit in the interpretations and presumptions of many non-economists, even though the scholars may lack well-articulated notions of a "production function."

We stress that the economic theory of production, technology, and technological advance, involves a set of rather particular assumptions. Because these assumptions seldom are made explicit, and because economists are so used to looking at input and output variation through the lens provided by the theory, in many minds the phenomena to be explained, and the theory used to provide explanation, blur together. These assumptions most definitely are not implicit in the tautology that outputs are to be explained by inputs and techniques employed, and that therefore differences in outputs can be explained in terms of differences in inputs and techniques. The economic theory of production, by stressing conscious choice, plays down the role of special circumstances, luck, or continuing experimentation, in determining actual inputs and techniques employed. The theory represses the role of individual differences in access to inputs or ability to use certain techniques. While the fact that output is a function of inputs and techniques employed may hold true as a tautology, the interpretation economists make of the observed relationships is very heavily dependent on the assumptions contained in the standard economic theory of production. An interpretation in terms of different rational and informed choices of generally available inputs and techniques is illuminating in some cases; but in other cases it blinds the observer to what actually lies behind the observed differences. Similarly, while technological advance may (tautologically) be defined as that which enables inputs to be employed more productively, it is not tautological that technological advance comes largely from an identifiable R&D activity.

In textbook microeconomics the economic activities used to illustrate the basic concepts of production theory are those like the production of steel, or aircraft, or corn, or the generation of electric power, or the shipping of oil through a pipe. For these it certainly makes sense, at least as a first approximation, to think of the principal inputs employed and techniques used as having come about largely as a result of managerial choice. The relevant inputs, and techniques, can to a first approximation be regarded as widely available. The technologies and contexts are sufficiently stable or predictable that it does not seem likely that much of observed activity should be regarded as an attempt to learn. The techniques themselves can be associated with an actual extant body of blueprints, designs, how to do it textbooks, and often the programs of professional training in the relevant "technologies." It also usually is possible to identify real organizations that actually do R&D, and to trace important advances in technological capabilities to these.

However, even for activities like the manufacture of steel, it is important to keep in mind that the concepts form a theory, which in some applications may be incomplete or even misleading. The quantity and quality of labor input may depend on the nature of management-labor relations, and the social customs of the plant, and these may be difficult to "choose" or control. Even steel technology is not tightly "blueprinted;" experience with technical assistance suggests that it is often very hard for one company to copy exactly what another company is doing even if the former has ample assistance from the latter. Also, many studies of advances in production processes in manufacturing show a considerable amount of continuing learning by doing, with only a limited role for separate R&D. Nonetheless, the orthodox theory of input-output relations and innovation has proved a useful first approximation in these traditional areas of application, at least for some purposes.

But the concepts have also been employed in recent years for activities such as crime prevention (Ehrlich, 1975; Witte, 1980), firefighting (Getz, 1979), the provision of medical services (Over and Smith, 1980), the raising of children (Leibowitz, 1977), and education. This article is concerned with the extent to which the orthodox ideas about input-output relations and innovation can be extended to these kinds of activities. The particular focus will be on education.

A number of studies have examined relationships between inputs and outputs in education and have attempted to interpret these as different points along a production function. But to what extent is it meaningful to regard these input-output relations in terms of a production function, as that concept is conventionally employed? Much of the research in question has tried to identify effective input packages or techniques used in certain schools or classrooms, with the presumtion that these could be adopted by those that are not now using them. But to what extent is it legitimate to think of what lies behind input-output relations in education as something like technological knowledge that can be shared among schools? A much touted response of the federal government to a widespread belief that education, even at the frontiers, isn't as good or efficient as it could be, has been to establish and fund special organizations to do educational R&D. But is it reasonable to believe that innovation in education can be created largely in specialized R&D institutions?

II. STUDIES OF SCHOOL EFFECTIVENESS

In this section we review two bodies of research and analysis concerned with school output and efficiency. The first, done mostly but not exclusively by economists, is concerned with finding resources that are associated with good performance by school systems, or schools or classes. The second, a tradition of research largely associated with schools of education, is concerned with identifying educational practices that will enhance effectiveness. While there are significant differences in methodology--in particular the research by scholars outside of economics is not guided explicitly by the idea of a production function--both bodies of research rest on the notion that there is an imitable "technology" involved in education, in that it is presumed that if one system or school or class can do something with certain effects, so can others. We shall question this premise at the conclusion of this section.

Quantitative Studies of School Effectiveness

Over the last fifteen years a large number of quantitative studies of the relationships between school resources and student achievement have been conducted. While there has been considerable variation in the characteristics of data bases and model specifications, these studies share a basic methodology. In particular, no attempt is made to manipulate experimentally the school resources children receive. Instead, "natural experiments"--the variation in school resources and other practices created by the operation of a school system--provide the data base for analysis. Multiple regression techniques are used to identify school resources and characteristics that are associated with effectiveness, and to assess the quantitative importance of these variables.

What has been learned from these kinds of studies of school effectiveness? Perhaps the most notable finding is that there are indeed significant differences in the amount of learning taking place in different schools and in different classrooms within the same schools (Armor et al., 1976; Hanushek, 1971; Murnane, 1975). Clearly schools differ, and the differences matter.

To what extent to differences in resources lie behind these differences in school effectiveness? Various studies report positive relationships between total expenditures per pupil and two types of educational outcomes--student achievement test scores (Bowles, 1968; Kiesling, 1967; Ribich, 1968; Winkler, 1975) and wage rates or incomes of graduates later in life (Johnson and Stafford, 1973; Morgan and Sirageldin, 1968; Wachtel, 1975; Welch, 1966). Greater resources applied thus seem to lead to enhanced educational effectiveness.

Other studies have examined the relationships between certain specific school resources and student achievement. It might be hoped that these studies would have thrown light on the questions of why dollars matter, and which kind of educational resources matter most. The results have been disappointing, however, and to some extent puzzling. Within the range of observation readily purchasable physical facilities and resources, of the sort commonly considered as being inputs to a conventional production function, are at best weakly and loosely related to student achievement. Thus, within the range of observed variation, the number of library books in the school, the quality of the science labs, the size and age of the school, in general are not statistically powerful determinants of school effectiveness. We will return later to this puzzle as to why total expenditure seems to matter, but particular kinds of expenditures do not.

On the other hand, certain characteristics of the humans--teachers and students--in the classroom matter considerably. Virtually every recent study of school effectiveness has found that some attributes of teachers are significantly related to student achievement. Intellectual skills of a teacher as measured by a verbal ability test matter (Coleman, 1966; Hanushek, 1971). The quality of the college (measured in various ways) the teacher attended tends to be significant (Summers and Wolfe, 1977; Winkler, 1975). Teachers with some experience tend to be more effective than teachers with no experience (Kean et al., 1979; Kiesling, 1981; Murnane, 1975; Murnane and Phillips, 1981a). Recent studies indicate that certain teachers are systematically more effective with certain types of students than with other types of students (Summers and Wolfe, 1977).

The student composition of the classroom also appears to matter. The evidence continues to hold up that elementary school children with low initial skill levels who attend schools in which the average achievement level is relatively high make more progress than such children who attend schools in which the average achievement level is relatively low (Henderson et al., 1978; Kean et al., 1979; Summers and Wolfe, 1977). There is similar evidence regarding socio-economic status (Winkler, 1975).

Surely these findings are intriguing, but how should they be interpreted? The interpretation in terms of a production function connotes that the left-hand side variable, effectiveness, is the consequence of the right-hand side variables, with at least some of the important latter under the control of the organization in question. As stressed in most cases when economists employ the production function interpretation of observed input-output relations, it is presumed that all right-hand side variables are objects of choice, and that what one organization does another could do. It is apparent that researchers doing the empirical work in education regard the relationship they are fitting as like a producting function. The results are viewed as pointing to what (all?) schools could do, if they had the knowledge and will. Indeed some economists have incorporated the results into optimizing algorithms designed to show how school districts should allocate resources to reach particular goals (Boardman, 1978; Cohn, 1980).

However, examination of the results creates significant doubts about this interpretation. The clearly defined physical inputs that are readily procurable by school districts don't seem to make much difference. The variables that do matter, teacher characteristics and the mix of students in the classroom, are not easy to "choose" in any fine grained way. Moreover, it is not known what resource allocation process resulted in the subtle resource configurations observed in the natural experiments. It may be that the resource allocation process itself influences the relationship between resources and student achievement (Murnane, 1981a; 1981b). This possibility, while long recognized by industrial psychologists, has only recently received significant attention by economists. However, recent work by Williamson (1975), Freeman and Medoff (1979), and others has served to increase awareness among economists of the possible effects that internal labor markets, labor unions and other institutions that influence the resource allocation process have on the relationship between human resources and outputs.

The conceptual awkwardness in applying the production function concept to education has been touched on, but only just touched on, by a number of economists. Levin (1976), for example, has posed the problem in terms of whether all schools can be assumed to operate at the technological frontier, or whether one must recognize that many schools are working within the frontier. Hanushek (1976) responded by proposing that even if the latter were true, as he had no doubt it was, it still made sense to try to explain these differences in effectiveness by introducing new variables into a regression format. Subsequent research with new variables has produced interesting results. However, the type of variables that have been added to recent multivariate studies of school effectiveness make the interpretation of the fitted relationship as like a production function even more problematic.

One variable included in recent studies is information on time use. At first glance, this does not seem to contravene the production function interpretation since time is a resource emphasized in much economics literature. The problem arises in the way time is measured. The early studies included variables such as the length of the school day or school year that were compatible with the production function notion in the sense that they seem to be variables at the discretion of some decision maker. However, weak relationships between these variables and student achievement led researchers to use more precise measures of time use. Recent work has focused on the amount of time that children actually spend "on task," working at particular subject matter. Data are collected by observers who measure, for example, how many minutes during an hour-long reading period students actually pay attention to the reading instruction. The research results do indicate that "time on task" is highly related to student learning (Brown and Saks, 1979). However, as anyone who has worked with children know, time on task is not a variable that can be easily manipulated.

Other variables included in recent studies include measures of teacher and student attitudes and quality of the classroom environment. Findings indicate that teachers who believe their students can learn are more effective than teachers who don't hold this belief (Link and Ratledge, 1979). Students who have relatively high self-esteem achieve more than other children (Boardman et al., 1977). The amount of learning that takes place in a classroom is negatively related to the frequency of disciplinary disruptions

(Hanushek, 1971).* These results are interesting and plausible. However, we don't know very much about how to create teachers with high expectations, students with high self-esteem and classrooms free of disciplinary problems. Nor do we know why certain schools have these characteristics and others do not. Thus, if it is these variables that matter, we must ask the basic question: is it reasonable to assume that what one school does another school could do if it had the same resources and the same "book of blueprints," where resources are defined as things that schools can purchase or manipulate and "book of blueprints" is treated as something readily communicable? Put another way, what choice does the school have regarding the values of the variables that seem to count? In the standard production function format the "right-hand side" variables are presumed to be a matter of choice. But "high teacher expectations for her students" does not look like a variable that one can go out and buy more of on the market, nor is it something any teacher can readily be taught to have.

Another approach to employing the production function concept to explain input-output relationships in education has been suggested by Brown and Saks (1975). These economists propose that the usual measures of school effectiveness, changes in student skill levels, should be viewed as the result of production function relationships and the decisions of teachers concerning the allocation of resources, especially their time, among students. In one respect this is an important advance over previous somewhat mechanistic thinking. The Brown and Saks approach does explicitly recognize that teachers do have discretion in allocating resources, and that to predict the outcomes of the schooling process, it is necessary to explain the factors that influence the behavior of the human resources.

---

* A causal interpretation of the findings of multivariate studies of school effectiveness is further cast into doubt by aberrant findings such as the finding reported by Summers and Wolfe (1977) of a positive relationship between the number of disruptions in the school and the achievement of students who started off with below grade level achievement.

In another respect, however, the Brown and Saks approach has proved disconcerting. To this point it has not been possible to separate the determinants of teachers' resource allocation decisions from production function relationships. Also, an obvious extension of the Brown and Saks thesis, and a thesis emphasized by many noneconomists (e.g., Comer, 1980; Sarason, 1971), is that students, as well as teachers, are decision makers in the classroom whose actions affect outcomes. This extension of the Brown and Saks model requires that school outcomes be viewed as the result of a production function and sets of preferences of students and teachers. In the context of such a complicated model the notion of a production function, with a well defined underlying technology that is separable from choice, loses operational significance. This brings us back to the dilemma posed earlier: how should we conceptualize relations between inputs and outputs in education?

Studies of Procedural Effectiveness

While studies by economists of factors influencing educational effectiveness started seriously less than twenty years ago, scholars in schools of education have been concerned with this question since the turn of the century. Callahan (1962) has argued that thinking about educational effectiveness has for a long time been influenced by analogies to manufacturing. Thus he quotes from a 1906 lecture by the dean of the School of Education at Stanford.

> Every manufacturing establishment that turns out a standard product or series of products of any kind maintains a force of efficiency experts to study methods of procedure and to measure and test the output of its works. Such men ultimately bring the manufacturing establishment large returns, by introducing improvements in processes and procedure, and in training the workmen to produce a larger and a better output. Our schools are, in a sense, factories in which the raw products (children) are to be shaped and fashioned into products to meet the various demands of life. The specifications for manufacturing come from the demands of twentieth century civilization, and it is the business of the school to build its pupils according to the specifications laid down. This demands good tools, specialized machinery, continuous measurement of production to see if it is according to specifications, the elimination of waste in manufacture, and a large variety in the output. (From Raymond E. Callahan, Education and the Cult of Efficiency, page 97.)

While economists studying school effectiveness have, until recently at least, tried to identify packages of inputs associated with strong performance, scholars in schools of education have focused on teaching techniques, curricula, and organizational design. By the early twentieth century experimental schools and other research activities were firmly established at the principal schools of education. A variety of activities and mechanisms were established to help schools find out about and adopt techniques that had been found superior in another setting (often a laboratory school).

Scholars in schools of education interested in "technology transfer," and economists studying input-output relations in education, look at education in quite different ways. Nonetheless, both tend to share the presumption that educational techniques are quite explicit and easily imitable. Experience has been no kinder to that presumption in technology transfer than to its application in production function fitting.

In particular, techniques or programs found to be successful in an original site have, with monotonous regularity, not had the same effect in other sites. Often performance in the original site has failed to be maintained. The proximate explanation for the inability to replicate successes is clear: the same curricula and instructional strategies are used in different ways in different sites. Van Deusen Lukas (1975) reports enormous variation in the actual educational practices taking place in classrooms using the same nominal instructional approach. Chall, in her well known book, Learning to Read: The Great Debate (1967), points out that even the basic distinction between the phonics approach to reading and the sight reading approach is not clear-cut when one observes their use in a number of classrooms.

The key question is what interpretation should be made of the variation in application and success of particular curricula and teaching techniques. Proponents of the technology transfer philosophy suggest two complementary interpretations. The first is that the problem lies in the lack of fidelity to the technical characteristics of the particular curriculum or instructional technique. Implicit in this view is the assumption that teaching and learning ought to proceed according to established, tested, best practice, and that fidelity to the details of a superior technology is possible and will result in increased productivity (cf. Benjamin, 1980). (It is interesting to note that Frederick Taylor, whose name is most commonly associated with these ideas for manufacturing industries, never tried to implement his ideas in an activity like teaching.)

The second interpretation, tightly related to the first, is that the discouraging research results stem from limitations in research methodology, including the difficulty of evaluating programs with non-experimental designs and of detecting possibly complex interactions between program characteristics and student characteristics. If this interpretation were correct, one could anticipate improvements in the ability to identify successful transferable educational programs since significant advances have been made in recent years in overcoming these methodological difficulties (e.g., Ashenfelter, 1978; Singer and Cohen, 1980). However, the application of these methodological advances requires that the set of possible "programs" (i.e., techniques and curricula) be well defined and that the appropriateness of the programs for particular students depend on readily observable characteristics of the students.

We do not believe that either of these complementary interpretations is correct. Instead, we suggest an alternative—that the variation in educational practice is unavoidable and in fact is crucial to effective teaching. The reason is that effective teaching requires information about the skills and personalities of students and about how students interact that only can be obtained during the classroom teaching process. Also, teachers vary greatly in the kinds of interactions they are good at. In other words, effective teaching requires intensive problem solving activity, and creative and personalized responses to frequent unpredicted circumstances. It is clear that this interpretation, which we believe is the correct one, casts a shadow on the faith that what one teacher or school is doing with success, another can replicate with comparable effect. It also casts a shadow on the research-limitation interpretation since our interpretation implies that neither of the critical assumptions implicit in using input-output research to identify successful replicable techniques is satisfied.

Let us step back for a moment and clarify what we are trying to argue. We are not arguing that the regression studies of the correlates of school effectiveness are not useful. The positive findings about the importance of teachers and peer groups should lead us to think about the factors that influence which teachers and which students end up in particular schools. Even the negative information that greater investment in facilities alone is unlikely to make a school much more effective is useful. Nor are we proposing that there are no systematic differences between teaching reading by sight and by phonics. There are differences in method; but the "blueprints" associated with different methods are quite sketchy. Many details have to be worked out by the individual teacher, and what one teacher does in applying a particular broadly defined method will diverge, often considerably, from what another teacher does. We are arguing that school authorities, while able to provide the context within which teachers go about their jobs, cannot control in any detail what a teacher does either through monitoring or through incentives. In thinking about the regression studies as estimating a "production function", and in thinking of teaching methods as tightly defined techniques, these facts get blocked from view.

III. R&D AND INNOVATION IN EDUCATION

As stated above, there is a long history of R&D in education. Emanating from these efforts, and from elsewhere, have come a flow of proposals for changing how education proceeds—new curricula, new methods, new ways of organizing classes and programs. In this section we consider first the structure and recent history of educational R&D, and the nature of the innovations that have come from it. We then turn to the various diagnoses that have been presented as to why educational R&D has not had more success.

Educational R&D and Innovation

Teaching, if it is to be done effectively, involves experimentation. This is inherent in the nature of the activity. Some children learn rapidly, others slowly; what is effective for one may not be effective for another. From time immemorial teachers have had to find out for themselves what works with which children and with which subject matter. There always have been teachers who have been particularly reflective about general principles and about particular techniques, and who have, as it were, systematically experimented.

Further, the tasks and constraints of schooling have not stayed static over the years. The objectives, clientele, and the context for education clearly were different in seventeenth century New England than they are in late twentieth century south Boston. Teachers, principals, and those responsible for larger administrative structures have had to adapt to change, and the adaptation surely has involved a considerable amount of trial and feedback experimentation. Put another way, educational innovation has been going on for centuries.

Conscious specialized educational R&D as a separately organized and sustained activity, has come onto the scene relatively recently, given this time sweep. By the turn of the century, however, it was well established, largely at schools of education which had as their mandate the training of teachers and school administrators, and increasingly the training of people who would do research and teach others how to do research. Some of the "R&D" pursued the goal of efficiency in a literal minded and somewhat pedantic way (recall the quote by the dean of the Stanford University School of Education). In some places, for example John Dewey's laboratory school (founded 1896) at the University of Chicago, educational R&D also meant the search for sounder general educational philosophy and practice.

The issue of the relative merits of phonics versus the visual methods of learning to read was, and continues to be, a subject of educational research. The question of how to measure achievement and the capabilities of students long has been of concern to scholars attempting to find ways to improve education. Testing techniques and measurement devices (e.g., I.Q. tests and achievement tests) have been an important part of the techniques that have been created through educational research and development.

From the beginning, educational R&D has been guided by the prevailing psychological theories bearing on how children learn. Educational psychology became an established field, located largely in schools of education, by the 1920s. The key ideas of the progressive education movement of the '20s and '30s—that students not only have more fun but learn better if they discover the answers for themselves rather than learning by rote, that the world itself is a marvelous laboratory for learning, etc.—emanated largely from philosophical and psychological thinking and research of scholars at schools of education.

Of course schools of education never did have a monopoly on educational

R&D. Psychologists in psychology departments long have been interested in how children learn. The research of scholars such as Piaget has had a profound effect on thinking about teaching. In recent years scholars concerned with how their subject was being taught in primary and secondary schools have involved themselves in curriculum design. The "new math", the "modern physics", and "modern language instruction" are striking cases where scholars of a field turned to the task of designing a curriculum for pre-college children.

The 1950s marks a watershed for educational R&D. Prior to that time federal involvement and funding were virtually nil. During the last quarter century the federal government has come to fund approximately 80 percent of work that is officially reported as educational R&D. The history of federal involvement evidences the changing goals, clientele, and context for education stressed above.

Federal involvement clearly began out of concern about the technical and scientific literacy of American primary and secondary school students, in a world where the Russians were beginning to look competent and dangerous technologically. Sputnik tweaked concerns that already were there, and the title of the National Defense Education Act of 1958 clearly signals the purpose of heightened Federal involvement in educational R&D. Even before that act, the National Science Foundation had begun to support the development of new science curricula. It was this NSF program that provided the support for the development of the new physics curriculum.

By the mid-1960s American concern about primary and secondary education had turned in a somewhat different direction--toward achieving greater equality of educational opportunity, and, in particular, toward improving the education of children considered to be disadvantaged. Title I of the

Elementary and Secondary Education Act of 1965 specifically funded the design of compensatory education programs for educationally disadvantaged children.

Until the mid-1960s, Federal support of educational R&D went largely to individuals and small groups of scholars, some of them at schools of education, some of them elsewhere. There was no attempt on the part of the federal government to design an organized system of educational R&D. Since the mid-1960s the Federal government has aimed for the building of such a system. Thus in 1964 the Federal government initiated support for the first of ten educational research centers. These centers were to be separate standing not-for-profit institutions, housed in universities. It was hoped that they would produce important breakthroughs in basic knowledge about education. In 1965, passage of the Elementary and Secondary Education Act provided funding for regional laboratories, the goal of which was to provide strong linkages between the results of basic research and the needs of classroom teachers. The drive to create a well organized, tidy system of educational R&D, and the importance of the National Institutes of Health as a mental model, is revealed clearly in the following quote from an influential paper prepared for the U.S. House of Representatives Select Subcommittee on Education, entitled "Educational Research and Development in the Sixties: the Mixed Report Card,":

> "Deficiencies in national planning, management, support and evaluation are a continuing impediment to realization of the full potential of educational R&D. These shortcomings spring largely from the failure to place educational research and development in charge of an adequately funded agency at a level in the government hierarchy comparable to the National Science Foundation or the National Institute of Health" (p. 54-55).

In 1972 Congress authorized the creation of the National Institute of Education. The hope was that this agency would be effective in coordinating educational research and in permitting educational research to develop the type of Congressional support that medical research enjoyed.

In sum, over the past quarter century the Federal government has played a powerful role not only in funding but also in shaping the evolution of the system of educational research and development. Prior to the intensive Federal involvement, educational R&D at universities was intimately involved with teaching programs. There was not much in the way of a formal separate R&D structure. The Federal government has, over the years, built up such a structure.

What has come out of that structure? It is interesting and revealing to peruse lists of educational innovations that have been deemed worth noting by those concerned with educational R&D and with increased educational effectiveness. Table I presents brief descriptions of twenty educational innovations drawn, randomly, from the 327 innovations listed in the U.S. Department of Education's volume Educational Programs That Work (1980). All of the programs described in the volume went through some sort of evaluation. The Joint Dissemination Review Panel (a group of professionals brought together under the auspices of the Department of Education) found these evaluations sufficiently well done and persuasive to warrant listing the innovations in the volume. For four of these innovations, Table II presents the more complete description of the program, together with the evidence of effectiveness presented in the volume.

The early advocates of educational R&D had in mind that education, and presumably educational innovation, were or could be analogous to industry and industrial innovation. Table III lists twenty twentieth century inventions drawn from the list of one hundred significant ones compiled by Jewkes, Sawers, and Stillerman (1969). As we have seen, more recently proponents of educational R&D have drawn the analogy to bio-medical R&D. Table IV lists "milestones" in the history of diagnosis and treatment of tuberculosis.

Perhaps a comparison of lists is unfair. The Jewkes, Sawers, and Stillerman list represents a careful, if subjective, winnowing of seventy years of inventive activity. The list of "milestones" in knowledge and treatment of tuberculosis is also a selective culling from a long history. The Department of Education list is both less selective, and covers only relatively recent "inventions". We are aware of no list of "major" educational innovations comparable to the Jewkes, Sawers, and Stillerman list of major twentieth century inventions or to the tuberculosis milestones. We have in Table V drawn up a short list of our own, based on unsystematic discussions with people knowledgeable about important changes, or attempted changes, in education.

Two post-World War II developments that many people regard as important involve the increasing use of computers, and television, in education. What is striking, however, about educational use of these devices is how varied that use is. Different schools, and different teachers, employ these new tools in different ways, with differing degrees of success. The "innovations" should be regarded as significant additions to teachers' bags of tools, to be used creatively, rather than as new "techniques". Further, the principal technological developments—computers and television—were developed elsewhere, and the technologies were in widespread use elsewhere before they were adapted to education. The technological R&D was not done at educational R&D laboratories. Educational R&D was concerned with exploring educational uses.

The aforementioned development of physics, mathematics, and language curricula by scholars of the fields in question are often cited as significant developments. These innovations involved the articulation of a different (from then present practice) broad design and philosophy for teaching a subject, and the development of various pieces of teaching equipment and teacher aids—textbooks, model exams, designs for experiments, language tapes, etc. To our knowledge no systematic study has been done of the impact of these innovations. Our interviews with several school officials suggest that today, twenty years or more after the new curricula were created, few schools adhere to any of them tightly, and in many the mode of instruction seems closer to the older technique than to the new ones. It appears that a number of schools that adopted and tried to use the new curricula have slid away from its use. Nonetheless, the teachers and principals we interviewed indicated that some pieces of the new curricula are in use, and that, more generally, the philosophies of the new curricula and the experience of trying to implement portions of them have significantly influenced thinking about how the subjects should be taught.

Innovations like "open classrooms" represent another genre. What is involved here is a proposed change in the philosophy and format for teaching young children, with less emphasis on a tightly structured curriculum and regime, and more focus on stimulating and responding to the intellectual interests of the children. Some teachers and schools have employed child-focused relatively loosely structured teaching styles for many years. The more recent surge of widespread interest in this broad philosophy, however, is attributable to a group of educational reformers who objected to the rigid classroom atmosphere they thought they saw everywhere, and to the report on the experimental school at Summerhill. "Open classroom", however, defines an educational philosophy and a broad way of doing things, not a particular narrowly constrained design or program. Classrooms can be open, or structured, in different degrees, and openness itself can mean virtually anything from anarchy to subtle but effective controls. While we know of no systematic record, it is clear that at various places open classrooms, initially at least, were regarded as laboratory experiments. However, the experimental and developmental work on the open classroom concept appears to have little in common with the R&D done at Bell Laboratories, or at DuPont, to perfect the characteristics and the production processes of the transistor and nylon.

Hopes, Disappointments, Diagnoses

We have seen how, starting in the mid-1950s, the Federal government has changed educational R&D from a modest enterprise, housed largely in schools of education and associated with the training of teachers and school administrators. By 1980 there had been created a large neatly structured system with laboratories, centers, clearing houses, and a National Institute of Education. All this was done with the hope and expectation that an educational R&D system organized similarly to that of the National Institutes of Health would put out similar research findings and technological developments. For many who early had high hopes, the results clearly have been disappointing. For example, a 1976 report sponsored by the National Institute of Education concluded:

> Education R&D does not have an impact on practice in education that is comparable with, for example, the impact of biomedical research on health care. (1976 Databook: The Status of Education Education Research and Development in the United States.)

Why not? What is the source of the problem? Many analysts have offered explanations. While the details vary, several common themes are apparent.

The first is that the failure of federal programs to promote successful educational R&D is due to poor management. This theme plays a central role in Dershimer's (1976) analysis of the history of Federal funding, and also in Sproull, Weiner, and Wolf's (1978) analysis of NIE. Dershimer, in particular, frequently draws analogies to NIH, with the implicit assumption that management style is the main explanation for the difference in the perceived successes of NIE and NIH.

A second theme is that educational research is young and more work is needed to develop the strong science base that underlies, for example, medical research (cf. Fundamental Research and the Process of Education, a Report to the National Institute of Education by the National Academy of Sciences, 1977). This diagnosis, if accepted as valid, could point in either of two directions. In one direction the thrust would be to continue to bet on a special educational research establishment, but to shift the allocation of research more toward basic and less toward applied, at least for a while. Alternatively, it could be argued that, since the present knowledge base does not permit effective applied R&D on education, funding of the specialized educational R&D establishment should be cut back, and more funding should go into the underlying sciences, such as developmental and cognitive psychology.

A third theme is that there is a great need to find a few powerful success stories—programs that work and can be disseminated to many sites. Federal attempts to find such programs include the funding of the annual document from which we have drawn our list of educational innovations—Educational Programs

That Work (1980).[*] It is hard to find any "powerful replicable success stories" in that document. It is unclear whether the perceived need to find a few reflects a belief that, if a few could be found, these would provide a model for other endeavors, or whether the need mainly reflects a wish that educational R&D had been more successful.

In any case, the common thread in all these diagnoses and prescriptions is that they accept, almost without question, the appropriateness of an R&D theory of innovation. The critical assumptions of this theory are: 1) Separate R&D activity standing at some distance from production and operated by specialists is the basic source of new knowledge; 2) R&D creates new products or processes that can be described accurately with blueprints or sets of instructions and which, therefore, can be replicated and diffused from place to place. In the preceding section we argued the inappropriateness of the "blueprint" theory of techniques used in educational practice. This section has been concerned with raising questions about the appropriateness of the conventional economic theory of R&D for describing educational innovation.

As in the preceding section, it is important to make clear what we mean and don't mean. We do not mean that educational R&D is a worthless endeavor. We do mean that it is a mistake to think of educational R&D as like industrial or biomedical R&D. We think it a mistake to think of innovation in education exclusively, or even largely, as an activity conducted in specialized facilities by specialists in R&D. Some of that kind of activity may be useful. But much of educational innovation we think ought to be linked more closely to teaching itself, and to the training of teachers and school administrators.

---

[*] The Educational Programs That Work series is not the only Federal attempt to document successful educational programs. For another example, see McLaughlin's (1975, Chapter 5) description of the creation of the "It Works" series, a group of case studies describing successful Title I compensatory education programs.

Rather than creating "programs that work" educational R&D should be viewed as part of the problem solving, experimenting, evaluating, adapting to new contexts and goals, that always is going on in education. If past experience be a guide to the future, it is a mistake to think of educational R&D as providing ready to use new technologies to schools and teachers. Rather, educational R&D provides a flow of ideas, broadly defined methods, evidence about what is being tried out in different settings and about how well particular initiatives have worked in these settings, which enrich capabilities for the experimentation and problem solving that go on in individual school systems, schools, and classrooms. Thus the relevant question is not "how widespread is the use of the modern physics package, or the new math package, and what have been the effects of use of these packages on performance." Rather, one might ask "what are the ways in which the new math and the modern physics ideas have influenced what goes on in classrooms, and in what ways, and in what contexts have these individual innovations enabled teachers to teach and students to learn more effectively." Unfortunately, this is not the way the evaluation questions have been asked. By inappropriately bringing to educational R&D a theory appropriate to a few manufacturing industries and a few areas of biomedical research, scholars have blinded themselves to what really is going on.

IV. TOWARD A THEORY OF INPUT-OUTPUT RELATIONS IN SECTORS WITH POORLY ARTICULATED TECHNIQUES

It should now be clear that we think the standard economic theory of production and innovation does not provide a useful way of thinking about effectiveness and innovation in education. What goes on in the classroom at any time--in orthodox jargon the techniques and inputs being employed--are "rationally chosen" to only a limited degree. To a considerable extent they are the result of a history of interaction between teacher and students and among students that have established a classroom culture many facets of which are of no one's choosing, often of no one's liking. Each class has a chemistry which is, to some extent, unique. The principal can dictate the curriculum, and the teacher can plan his or her classroom procedure for a particular lesson and, to some limited extent, carry it out. However, much of what happens is under no one's detailed control. The characteristics of individual teachers influence both the strategies they choose, and what actually happens. But the strategies that one teacher feels comfortable with may be quite awkward for another to attempt, and what actually happens in the classroom inevitably involves some frustration as well as accomplishment for teacher as well as student. The curriculum and its scheduled presentation, and the teacher's broad gauged strategy and style, may remain doggedly consistent in the face of continuing partial frustration. However, classroom activity inevitably involves as well continuing problem solving, experimentation, and search for better ways to motivate or control a particular student or to get across a point that somehow students do not understand (Sarason, 1981). In a nutshell, in contrast with the implicit assumptions of orthodox production theory, in education "techniques" and "inputs" to a considerable extent evolve rather than are rationally chosen; the idiosyncratic elements are important as well as the standard ones; a good portion of input and energy is continually involved in problem solving as contrasted with smooth operation of a routine.

These characteristics help to resolve what otherwise could be regarded as a puzzle about the findings of research on educational effectiveness. As pointed out in Section II many early studies of the determinants of effective education found a positive relationship between per pupil expenditure levels and educational outcomes. However, when expenditures were decomposed so that educational outcomes could be regressed against specific resources, such as the number of books in the library, or the number of teachers per pupil, consistent significant relations were not found. That dollars matter, but particular resources do not, is something of a puzzle if education is viewed within the framework of orthodox production theory. But these findings are consistent with the view that considerable experimentation generally is needed all the time to find out what works and what doesn't. Specific resources do not consistently matter because students' needs and teachers' capabilities vary widely and the resources that work best will also vary—in ways that can only be discovered through a process of trial, error, and adaptation. Dollars matter because they improve the conditions under which experimentation takes place—by providing a wider range of materials with which teachers can experiment and adapt—and by facilitating the hiring and retention of teachers who are more efficient problem solvers.

Viewing teaching through the lens suggested above also throws light on a second puzzle—namely, the poor track record of the education industry in replicating "successful" innovations. What is actually done is only partly a matter of conscious choice. What will work and what will not work varies from situation to situation. Much of problem solving and fine tuning inevitably must go on in the particular classroom, and thus what someone else has done successfully can provide only gross guidance as to what might (or might not) be effective in a different context.

We find it interesting that sophisticated participants in successful educational innovations both understand this, and resist it. Thus, descriptions of successful innovations written up by the innovators themselves (e.g., Comer, 1980; "It works" series, 1969) tend to stress detailed description of the techniques, procedures, and materials that characterize the innovations in their final form. In other words, the descriptions emphasize the blueprints that resulted from the innovation process—blueprints that describe procedures that are effective only at that site at that time.

However, a careful reading of the prefaces, the footnotes, and sometimes the introductory chapters reveals an alternative view of the ingredients of a successful innovation--ingredients that may provide more lessons to potential imitators than the blueprints do. These ingredients are a sustained process of experimentation, guided only loosely by theory, and usually characterized by false starts, changes in direction, creative adaptation, and, most of all, by extraordinary persistence, often in the face of adverse circumstances.

We suggest that understanding more about the determinants of high quality education requires paying less attention to formulating and disseminating blueprints and more attention to examining why the process of experimentation proceeds in some settings with enthusiasm, skill, and persistence, while in other settings, experimentation and creative problem solving take place only very slowly and often ineptly.

While the examples in this paper are drawn from education, it is only one of many activities with poorly articulated techniques. Other relevant examples include crime prevention (see Ehrlich, 1975; Witte, 1980), and the provision of medical services (Over and Smith, 1980). In such activities, one would expect to see considerable variation at any time in "effectiveness" among economic units, variation that cannot be eliminated by better dissemination of blueprints. Will not selection pressures result in performance that is pretty much uniform? Not if, as in education, the environment is not strongly competitive. But even in sectors organized through markets, where techniques contain a considerable tacit element and individual skills differ significantly, the "fittest" may be the richest, but there may be considerable room for the less fit.

Poorly articulated techniques not only hinder cross organizational teaching and learning, they restrain the ability of an organization to replicate itself. Excellent, profitable restaurants show a lamentable tendency to become less excellent after they expand, set up a branch, hire some new chefs. It is interesting that restaurants like McDonald's and Howard Johnson's, which have managed to develop a system for replication, are marked by uniform practiced blueprinted mediocrity.

How does innovation come about in activities with poorly articulated techniques? What is the role of separate specialized R&D? Because techniques are tacit, innovation--the creation and introduction of a new technique--can only define a broad way of doing things, not something sharp and precise. Because separate R&D can't "work out the details," freestanding R&D is likely to be of smaller importance than in activities where separate R&D can work out the details as well as the broad designs. However, in such activities separate R&D can still play a role if these are special competences and knowledge that are useful in identifying, designing, and carrying out experimental departures from prevailing practice. But it does not play the role of generating well articulated new blueprints; rather, it provides an intellectual ambiance, as it were, within which organizations strike out in new directions.

Again, one last time, let us make our position clear. We are not arguing that what we have called the standard theory of input-output relations, and innovation, should be scrapped. Rather, we want to draw attention to the fact that it is a theory, not a simple description of empirical reality, and that its range of applicability appears to be limited. It has blinded, not illuminated, research on input-output relations and innovation in education, we argue. And education is not unique. There are many activities with similar characteristics. In this concluding section we have offered some thoughts on what a theory of input-output relations, and innovation, relevant to these activities might look like.

TABLE I

A Selection From Educational Programs That Work

Project	Description
SECONDARY CREDIT EXCHANGE PROGRAM	A continuation school for secondary grade migrant students who have been attending school in another district or state and are not able to continue school because of the need to work.
CAREER EDUCATION RESOURCE CENTER PROGRAM (CERCP)	A program to increase self-appraisal, occupational information, goal selection, and planning and problem solving among low-achieving urban tenth graders.
PROJECT CDCC: Career Development Centered Curriculum	A sequentially coordinated career development program with teacher-developed classroom units that focus on career awareness, self-awareness, and introduction to decision making.
ADDED DIMENSIONS TO PARENT AND PRESCHOOL EDUCATION	An early childhood education program built on the belief that parents are their child's first and most effective teachers.
THE CHILD DEVELOPMENT CENTER	A model community center designed to identify children's special developmental needs and provide appropriate interventions before school begins, through joint cooperation of parents, kindgarten teachers, and special educators.
COPE: Cognitively Oriented Pre-Primary Experience	A comprehensive sequentially programmed preprimary curriculum and management system that provides for individual developmental growth and learning of basic readiness skills.
OAKLAND UNIFIED SCHOOL DISTRICT FOLLOW THROUGH PROGRAM: Learning Through Literature	An interdisciplinary approach for students in grades K-2 utilizing a literature-centered curriculum.
ISIS: Individualized Science Instructional System Dissemination Project	An interdisciplinary, modular science program preparing students who do not plan to major in postsecondary science to understand practical, real-world, science-related problems.
COMPUTER-BASED PLANNING RESOURCES: Project Simu-School	A program using computer-based resources for improved educational planning and decision making.

TABLE I (continued)

Project	Description
BASIC: Basic Adaptable Skills for the Individual Child	Four separate but interrelated programs consisting of highly structured, sequential, individualized curricula in readiness, mathematics, and reading.
COMPREHENSIVE SCHOOL MATHEMATICS PROGRAM (CSMP)	An exciting, complete elementary-level mathematics curriculum from basics to problem-solving for students of all ability levels.
EARLY CHILDHOOD PREVENTIVE CURRICULUM (ECPC)	A program for high-risk first-grade students developing the perceptual, cognitive, and language skills they need to respond successfully to beginning reading instruction.
METRICS MADE EASY	A systematic hands-on approach to metrication that includes field-tested techniques, activities, an assessment system, and staff development materials.
READING ENGLISH ROTATION PROJECT	A rotating classroom approach to teaching reading skills to students, grades 7-9. Not a pull-out model.
READING — INDIVIDUALIZED REMEDIAL LABORATORIES MATH — INDIVIDUALIZED REMEDIATION	A project designed to provide continuous diagnosis of student needs and daily prescriptions for learning improvement.
INTERACTIVE CURRICULAR EXPERIENCE	A process approach to developing individualized programs for the handicapped student using home, school, and community resources.
MODIFICATION OF CHILDREN'S ORAL LANGUAGE	A special program for training staff to work with students having language disabilities.
PROJECT SHARE: Sharing High Yield Accountability with Resource Educators	An instructional process for remediation of basic skills in learning-disabled students in mainstream education.
INSTITUTE FOR CREATIVE EDUCATION	A sequentially ordered curriculum that teaches a creative problem-solving process using tasks linked to a wide variety of subject areas.
OMBUDSMAN	A school-based semester-long drug education/ primary prevention program.

TABLE II

Details on a Sample from Table I

Project: THE CHILD DEVELOPMENT CENTER

Description: A model community center designed to identify children's special developmental needs and provide appropriate interventions before school begins, through joint cooperation of parents, kindergarten teachers, and special educators.

The Child Development Center is based on the conviction that the sooner educators identify young children's developmental needs and work together with parents to achieve effective interventions, the stronger the chance of children's early success in school. This program offers an ongoing format of child/parent services starting the September before the child is age-eligible for kindergarten. This format includes a diagnostic center, where preschoolers receive evaluations from a school nurse, speech/language specialist, and psychologist; a parent resource center, where parents meet with educators to discuss early-childhood topics and concerns; a child-study center, where preschoolers meet periodically with peers for group interaction while being observed by parents and center staff; and a referral center, which provides children having special developmental needs with interventions before school begins and which transmits information on all program children to parents and kindergarten teachers.

The program supplies a motivational slide-tape presentation for parents and educators, a parent handbook containing child-development articles and materials, a teacher's manual containing instructions and materials for implementing the program, and an end-of-year program booklet for recording and relaying developmental information about the child to parents, kindergarten teachers, and special educators.

Evidence of Effectiveness: In 1976, experimental children achieved significantly higher posttest scores than control-group counterparts on the McCarthy Scales of Children's Abilities (Verbal, Perceptual, General Cognitive, Memory, and Motor subtests) and the Goldman-Fristoe Tests of Articulation.

Project: BASIC: Basic Adaptable Skills for the Individual Child

Description: Four separate but interrelated programs consisting of highly structured, sequential, individualized curricula in readiness, mathematics, and reading.

Individually Prescribed Instruction (IPI) curricula were developed at the University of Pittsburgh's Learning Research and Development Center. The IPI program was designed to meet the individual developmental needs of young children in the following

TABLE II (continued)

basic areas: Primary Education Program and Perceptual Skills, which emphasizes pre-reading and math objectives in a hierarchical order for preschool, kindergarten, and special education students in quantification, classification, visual motor, auditory motor, general motor, and letters and numerals; Early and Primary Reading, a phonetic approach emphasizing mastery of specific reading skills; Math, stressing individual progress with continuous growth in numeration/place value, addition, subtraction, multiplication, division, fractions, money, time, systems of measurement, geometry, and application; Intermediate Reading, stressing individual and group instruction in comprehension skills.

Each BASIC component emphasizes student-management skills, positive reinforcement, continuous testing, accurate and well-defined record keeping, and parent involvement. The curriculum is characterized by five critical elements: structured curricula for each content area comprised of a series of behavioral objectives arranged in a hierarchical order by unit and level; an assessment system comprised of criterion-referenced tests matched to curriculum objectives; a management system designed to provide individual programs and learning experiences for children; individualized instructional materials, sponsor-developed commercial sources, and teacher-constructed materials; and a monitoring and record-keeping system depicting the location and mastery level of each student in each curriculum area.

Evidence of Effectiveness: Participation in BASIC results in increased levels of achievement in reading and math.

Project: READING — INDIVIDUALIZED REMEDIAL LABORATORIES
MATH — INDIVIDUALIZED REMEDIATION

Description: A project designed to provide continuous diagnosis of student needs and daily prescriptions for learning improvement.

The reading laboratories have been developed for high concentration on the improvement of basic reading skills. A reading laboratory staffed by one special reading teacher and a paraprofessional accomodates 80-120 students daily for the entire school year. Each student's daily prescription includes two or more activities designed to meet his/her needs. Students' prescriptions include programmed and self-instructional materials purchased from a variety of vendors or developed both by consultants and project teachers. Emphasis is placed on inservice education, focusing on cognitive reading skills and on the management and use of individualized instruction in the classroom. Inservice education is provided through workshops, consultant classroom visits, and local supervisory services and support.

The mathematics program provides systematic remedial instruction in areas of individual student weakness. A teacher and a paraprofessional work with 80-120 students daily in a specially equipped classroom. The mathematics laboratories are characterized by a focus on carefully selected essential concepts, skills, and

TABLE II (continued)

applications with number ideas and computation; an individualized approach to the instruction; a meaningful approach to the learning of content; careful monitoring of student achievement; and teacher guidance in a supportive atmosphere. The program is based on project-developed materials, reinforced by a variety of supplementary resources and activities. Daily work is guided by individual prescriptions consisting of two or three types of activities. Inservice education receives a strong emphasis through workshops, consultant visits, and local supervisory services and support. There is an ongoing evaluation of project content, materials, instructional procedures, and overall achievement pattern of students.

Evidence of Effectiveness: Data collected in 1977-78 following JDRP approval showed the following average gains over an eight-month period: Slosson Oral Reading Test, 1.28 years average gain; Gray Oral Test, 1.51 years average gain; California Reading Achievement Test, 1.15 years average gain; California Math Test, 1.11 years average gain. The individualized laboratory approach seems also to minimize student negative behavior through individual and positive reinforcement.

Project: OMBUDSMAN

Description: A school-based semester-long drug education/primary prevention program.

Ombudsman is a structured course designed to reduce certain psychological and attitudinal states closely related to drug use. Ombudsman does not emphasize information about drugs per se, although some drug topics are included for discussion as part of specific exercises.

The course has three major phases. The first phase focuses on self-awareness, and includes a series of exercises permitting students to gain a wider understanding and appreciation of their values as autonomous individuals. The second phase teaches group skills and provides students with an opportunity to develop communication, decision-making, and problem-solving techniques that can be applied in the immediate class situation as well as in other important group contexts such as family and peers. The third, active, phase is in many ways the most important: the class uses the insights and skills gained during the first two phases to plan and carry out a project within the community or school system. During this phase, students have an opportunity to experience the excitement and satisfaction of reaching out to others in a creative and constructive way.

The program must be presented to a given classroom of students for a minimum of two hours per week for a full semester.

Evidence of Effectiveness: Pre- and posttesting of experimental and control groups (1977) illustrated Ombudsman's impact on a series of high-risk states related to drug use. Longitudinal comparisons of Ombudsman graduates and non-Ombudsman students (1977) have demonstrated that program graduates are more likely to give up drug use.

TABLE III

A Selection of Twentieth Century Inventions

- Automatic Transmission
- Ball-Point Pen
- 'Cellophane'
- Continuous Casting of Steel
- Cyclotron
- Fluorescent Lighting
- Helicopter
- Radio
- Television
- Transistor
- Xerography
- Zip Fastener
- Air Cushion Vehicles
- Oxygen Steel-making
- Electronic Digital Computers
- Jet Engine
- Kodachrome
- Magnetic Recording
- Polyethylene
- Photo-Typesetting

SOURCE: John Jewkes, David Sawers, and Richard Stillerman, *The Sources of Invention*, 2nd ed., (New York: W.W. Norton & Co.), pp. 231-356, 1969.

TABLE IV

Milestones in the History of Tuberculosis

460-370 B.C.	Hippocrates provides an accurate clinical description of tuberculosis.
1679	Sylvius sees tubercles (nodules in the lungs) as the precursors of tuberculosis.
1810	Bayle teaches that tubercles are a specific local formation causing a specific disease.
1819	Laennec establishes the unity of the tubercle. He invents the stethoscope and founds modern physical diagnosis.
1859	Brehmer begins the modern institutional treatment of tuberculosis.
1865	Villemin demonstrates experimentally that tuberculosis is a specific infection due to an inoculable agent.
1882	Koch discovers the tubercle bacillus, the infectious agent of tuberculosis.
1890	Koch produces tuberculin, a glycerine extract of tubercle bacilli.
1894	Forlanini pioneers the use of artificial pneumothorax in the surgical treatment of pulmonary tuberculosis.
1895	Roentgen discovers x-rays.
1907	Von Pirquet introduces a simple tuberculin skin test.
1936	Mass chest x-ray survey methods introduced.
1944	Waksman and associates discover streptomycin, the first effective anti-tuberculous drug.

SOURCE: Richard M. Burke, An Historical Chronology of Tuberculosis, 2nd ed., (Springfield, Illinois: Charles C. Thomas), pp. 77-78, 1955.

TABLE V

Significant (Our Judgement) Innovations in Education

Class Use of Computers

Class Use of Television

Modern Physics Instruction

New Math Instruction

Modern Language Instruction

Open Classroom methods

REFERENCES

Armor, David et al. Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools. Santa Monica: The Rand Corporation, 1976.

Ashenfelter, Orley. "Estimating the effect of training programs on earnings." Review of Economics and Statistics, February 1978, 60, 47-57.

Benjamin, Robert. Making Schools Work. New York: Continuum, 1980.

Boardman, Anthony E. "Policy Models for the Management of Student Achievement and Other Educational Outputs." Management Science, 1978, 8, 99-122.

Boardman, Anthony E., Otto A. Davis and Peggy R. Sanday. "A Simultaneous Equations Model of the Education Process." Journal of Public Economics, 1977, 7.

Bowles, Samuel S. "Toward Equality of Educational Opportunity." Harvard Educational Review, Winter 1968, 38.

Brown, Byron W. and Daniel H. Saks. "The Allocation and Use of Instructional Time in Reading and Mathematics: Some Theory and Evidence." 1979. Paper presented at the 1980 Meetings of the American Educational Research Association, Boston, Mass., April 1980.

________. "The Production and Distribution of Cognitive Skills Within Schools." Journal of Political Economy, 1975, 83, No. 3, 571-593.

Burke, Richard M. An Historical Chronology of Tuberculosis, 2nd edition. Springfield, Ill.: Charles C. Thomas, 1955.

Callahan, Raymond E. Education and the Cult of Efficiency. Chicago: University of Chicago Press, 1962.

Chall, Jeanne. Learning to Read: The Great Debate. New York: McGraw-Hill, 1967.

Chase, Francis S. "Educational Research and Development in the Sixties: The Mixed Report Card." A background paper submitted to the Select Subcommittee on Education, U.S. House of Representatives, April 1971.

Cohn, Elchanan. "Combining Efficiency and Equity: Optimization of Resource Allocation in State School Systems" in McMahon, Walter W. and Terry G. Geske (eds.), *Toward Efficiency and Equity in Educational Finance*. Boston: Allyn and Bacon, 1980.

Coleman, James S. et al. *Equality of Educational Opportunity*. Washington, D.C.: Office of Education, National Center for Education Statistics, 1966.

Comer, James P. *School Power. Implications of an Intervention Project*. New York: Free Press, 1980.

1976 Databook. The Status of Education Research and Development in the United States. Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education and Welfare.

Dershimer, Richard A. *The Federal Government and Educational R & D*. Lexington, Mass.: Lexington Books, 1976.

Educational Programs That Work, 7th edition, Fall 1980. Prepared for the National Diffusion Network, Division of Educational Replication, Department of Education by Far West Laboratory for Educational Research and Development.

Ehrlich, Isaac. "The Deterrent Effect of Capital Punishment: A Question of Life and Death." *American Economic Review*, June 1975, *65*, 397-415.

Freeman, R.B. and J.L. Medoff. "The Two Faces of Unionism." *The Public Interest*, 1979, *57*.

Getz, Malcolm. *The Economics of the Urban Fire Department*. Baltimore: The Johns Hopkins University Press, 1979.

Hanushek, Eric. "Comments" in Froomkin, Joseph T., Dean T. Jamison and Roy Radner (eds.), Education as an Industry. Cambridge: Ballinger Publ. Co. for the National Bureau of Economic Research, Inc., 1976.

________. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." American Economic Review, 1971, 61, 280-288.

Henderson, V., P. Mieszkowski and Y. Sauvageau. "Peer Group Effects and Educational Production Functions." Journal of Public Economics, 1978, 10, 97-106.

It Works series. Oakland, California, Preschool Program. Preschool Programs in Compensatory Education I, Office of Education, 1969 (OE 37057).

Jewkes, John, David Sawers, and Richard Stillerman. The Sources of Invention, 2nd edition. New York: W.W. Norton, 1969, 231-356.

Johnson, G.E. and F.P. Stafford. "Social Returns to Quantity and Quality of Schooling." The Journal of Human Resources, Spring 1973.

Kean, M.N., A. Summers, J.J. Raivetz, and I.J. Farber. "What Works in Reading?" and "What Works in Reading?--Technical Supplement." Office of Research and Evaluation, School District of Philadlephia, 1979.

Kiesler, Sara B. and Charles F. Turner (eds.). Fundamental Research and the Process of Education. Washington, D.C., 1977. Final Report to the National Institute of Education by the Committee on Fundamental Research Relevant to Education, Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences.

Kiesling, Herb J. "The Production of Reading." Indiana University, mimeo. 1981.

________. "Measuring a Local Government Service: A Study of School Districts in New York State." Review of Economics and Statistics, August 1967, 49.

Leibowitz, Arleen. "Parental Inputs and Children's Achievement." The Journal of Human Resources, Spring 1977, 12, No. 2.

Levin, Henry M. "Concepts of Economic Efficiency and Educational Production" in Froomkin, Joseph T., Dean T. Jamison and Roy Radner (eds.), Education as an Industry. Cambridge: Ballinger Publishing Co., for the National Bureau of Economic Research, Inc., 1976.

Link, Charles R. and Edward C. Ratledge. "Student Perceptions, IQ, and Achievement." The Journal of Human Resources, Winter 1979, 14, No. 1.

McLaughlin, Milbrey Wallin. Evaluation and Reform. The Elementary and Secondary Education Act of 1965, Title I. Cambridge: Ballinger, 1975.

Morgan, J.N. and I. Sirageldin. "A Note on the Returns to Quality of Schooling." Journal of Political Economy, September/October 1968.

Murnane, Richard J. "Interpreting the Evidence on School Effectiveness." Teachers College Record, Fall 1981a, 83, 19-35.

________. "Seniority Rules and Educational Productivity: Understanding the Consequences of a Mandate for Equality." American Journal of Education, November 1981, 90, 14-38.

________. The Impact of School Resources on the Learning of Inner City Children. Cambridge: Ballinger, 1975.

Murnane, Richard J. and Barbara Phillips. "Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance." Economics of Education Review, 1981a (forthcoming).

________. "What Do Effective Teachers of Inner City Children Have in Common?" Social Science Research, 1981b, 10, 83-100.

Nelson, Richard R. "Production Sets, Technological Knowledge, and R & D: Fragile and Overworked Constructs for Analysis of Productivity." American Economic Review Papers and Proceedings, May 1980, 62-67.

Nelson, Richard R. and Sidney Winter. An Evolutionary Theory of Economic Capabilities and Behavior. Cambridge: Harvard U. Press, forthcoming, 1982.

________. "Neoclassical vs. Evolutionary Theories of Economic Growth: Critique and Prospectus." The Economic Journal, December 1974, 886-905.

Over, A. Mead, Jr. and Kenneth R. Smith. "The Estimation of the Ambulatory Medical Care Technology Where Output is an Unobservable Variable." Journal of Econometrics, 1980, 13, 225-251.

Ribich, Thomas I. Education and Poverty. Washington, D.C.: The Brookings Institution, 1968.

Sarason, Seymour B. The Culture of the School and the Problem of Change. Boston: Allyn and Bacon, 1971.

Singer, Burton and Joel E. Cohen. "Estimating Malaria Incidence and Recovery Rates from Panel Surveys. Mathematical Biosciences, 1980, 49, 273-305.

Sproull, Lee, Stephen Weiner, and David Wolf. Organizing an Anarchy. Belief, Bureaucracy, and Politics in the National Institute of Education. Chicago: U. of Chicago Press, 1978.

Summers, Anita A. and Barbara L. Wolfe. "Do Schools Make a Difference?" American Economic Review, September 1977, 67, No. 4.

Van Deusen Lukas, Carol. "Problems in Implementing Head Start Planned Variation Models" in Rivlin, A. and P.M. Timpane (eds.), Planned Variation in Education. Washington, D.C.: The Brookings Institution, 1975.

Wachtel, Paul. "The Effect of School Quality on Achievement, Attainment Levels, and Lifetime Earnings." Explorations in Economic Research, Fall 1975, 2.

Welch, F. "Measurement of the Quality of Schooling." American Economic Review, May 1966.

Williamson, O.E. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press, 1975.

Winkler, D.R. "Educational Achievement and School Peer Group Composition." *The Journal of Human Resources*, 1975, *10*, 189-204.

Witte, Anne Dryden. "Estimating the Economic Model of Crime with Individual Data." *Quarterly Journal of Economics*, February 1980, *94*, No. 1, 57-84.

Seniority Rules and Educational Productivity:
Understanding the Consequences of a Mandate for Equality

by

Richard J. Murnane

Department of Economics
and
Institution for Social and Policy Studies
Yale University

August 27, 1981

The author would like to thank Samuel Brownell, James Fesler, Charles Lindblom, Richard Nelson, Sharon Oster, Edward Pauly, Seymour Sarason and John Simon for helpful comments on an earlier draft.

This paper is based on research funded in part by the National Institute of Education under Grant NIE-G-79-0084, and in part by the Institute for Research on Educational Finance and Governance, Stanford University.

ABSTRACT

This paper analyzes the impact on the performance of public education of seniority rules in teachers' contracts. Part I considers the relative merits of alternative types of employment contracts for teachers in light of recent economic literature on the determinants of efficient contracts and recent empirical work on the production process in education. The key point is that, given the technological characteristics of the education process, the mandate to provide an education to all children, and the lack of low cost educational options available to low income families, seniority-based employment contracts may be more effective in promoting public education than performance-based contracts.

Part II discusses the impact of seniority rules on the performance of public education during the last ten years, a period characterized by declining enrollments and growing power of teachers' unions. This part points out that seniority rules have created significant problems for school districts coping with declining enrollments. However, many of these problems stem from the lack of expertise on the part of school administrators and leaders of teachers' unions. The analysis suggests that as these agents have gained expertise, they have been able to find solutions to the problems of declining enrollments that mitigate the deleterious consequences of seniority rules, while retaining their positive contributions.

INTRODUCTION

Many Americans are unhappy with the quality of public education in the United States. Costs continue to rise, even though the number of students has fallen in recent years, and indicators such as SAT scores and minimum competency test results show that many children are leaving the public schools without adequate preparation for college work and even without basic skills. Much of the criticism of the public schools in recent years has focused on personnel policies for teachers. One reason is that teachers' salaries and fringe benefits account for 70 to 80 percent of the current account budgets of most school districts. A second reason is that research evidence indicates that teachers are the school resource most important in determining how much children learn in school (Hanushek, 1979).

The purpose of this paper is to analyze the impact on the performance of public education of one critical aspect of personnel policy, namely, the role of seniority rules in determining the salaries and job security of teachers. The paper has two parts. Part I considers the relative merits of alternative types of employment contracts for teachers in light of recent economic literature on the determinants of efficient contracts and in light of recent empirical work on the nature of the production process in education. The key point of this part is that, given the technological characteristics of the education process, the mandate to provide an education to all children, and the lack of low cost educational options available to low income families, seniority-based employment contracts may be more effective in promoting education than contracts that base teachers' compensation and job security on assessments of their performance.

Part II of the essay discusses the impact of seniority rules on the performance of public education during the last ten years, a period characterized by declining enrollments and growing power of teachers' unions. This part points out that seniority rules have created significant problems for school districts coping with declining enrollments. However, many of these problems stem not from the rules themselves but rather from a lack of expertise on the part of school district administrators and leaders of teachers' unions, the agents responsible for negotiating and administering interpretations of these rules. The analysis suggests that as these agents have gained expertise in consultation and negotiation, they have often been able to find solutions to the problems of declining enrollments that mitigate the deleterious consequences of the seniority rules, while retaining their positive contributions.

I. CONTRACT TYPES AND TEACHER RESPONSES

A. SENIORITY RULES FOR TEACHERS

In most school districts in this country, the salaries of public school teachers are determined by a unified salary schedule applying to all teachers in the district. In the schedules in use in most districts, the salary of an individual teacher is determined exclusively by the number of years the teacher has taught and by the highest degree the teacher has earned. Evaluations of teacher performance rarely have an impact on salaries.

Typically, job security is also determined primarily by seniority. This has become an important issue in recent years as budget stringency and declining enrollments have reduced the demand for teachers and necessitated transfers and layoffs of large numbers of teachers. In most districts the rules governing transfers and layoffs contain the following steps:

1. When declining enrollments necessitate a reduction in the teaching staff of a school, the teacher in that school with the least seniority loses his or her position.
2. This teacher may then transfer to any vacant position in the school system for which he or she is qualified.
3. If no vacancy exists, this teacher may displace the teacher in that particular job category who has the least seniority in the system.

The common thread running through the rules is the primary role of teacher seniority.

A criticism often made of contracts that place heavy reliance on seniority rules is that they provide teachers with perverse incentives, and as a result, make the system less efficient. The key points in the argument are that ineffective teachers are given no clearcut signals that their performance must be improved; effective teachers are given no special rewards. As a result, poor teachers remain in the profession without improving their performance, while talented teachers, discouraged by the lack of rewards for effective teaching and attracted to professions in which salaries are related to productivity, leave the profession.

One implication of this criticism is that the delivery of educational services would be more efficient if performance were the primary determinant of the compensation and job security of teachers. This implication is valid if it is possible with reasonable monitoring costs to assess the performance of individual teachers accurately. Recent contributions to economic theory have clarified the circumstances under which this condition is fulfilled.

B. EFFICIENT CONTRACTS

The economic literature on employment contracts addresses the following question: What factors influence the efficiency of different types of employment contracts? This section draws from this literature to describe the factors influencing the efficiency of three types of employment contracts. These descriptions will prove useful in analyzing employment contracts for public school teachers.

Consider first employment contracts in which the compensation and job security of the individual worker are based on a measure of the worker's output. The simplest of such contracts is the piece rate contract. Such a contract is efficient if the contribution of the individual worker to the firm can be measured accurately at relatively low cost. Commercial laundries' contracts with workers who iron shirts provide an example. A single worker does the entire ironing of any given shirt so the problem of joint products is not present. Counting the number of shirts ironed is inexpensive, and the problem of poor quality is controlled by customer complaints.

A second type of employment contract specifies that compensation and job security depend on supervisors' assessments of observed actions of individual workers. This type of contract is common in situations in which employees work in groups and the value added of individual workers cannot be determined, but their contribution can be assessed by their effort level and by the extent to which they adhere to actions known to be related to productivity. Alchian and Demsetz (1972) provide the example of workers employed to unload a truck. Since several laborers work as a team unloading a truck, the output of an individual worker cannot be measured. However, since the actions required to empty the truck are well understood and easily observable, the performance of individual workers can be assessed by observing their actions. These assessments can then be used to dismiss incompetent workers and to reward exceptionally productive workers.

The requirements for this type of contract to be efficient are that the relationships between worker actions and desired output be clear-cut and that the costs of monitoring worker actions be low relative to the productivity gains associated with an incentive system that bases compensation and job security on assessments of performance.

A third type of employment contract specifies that compensation and job security are determined by internal labor market rules. While the precise details of these rules vary, typically seniority plays a dominant role. As Williamson (1975) has explained, this type of contract is efficient in work situations characterized by the following two conditions:

1. As a result of on-the-job experience, individual workers acquire specialized knowledge, the use of which has a significant impact on the performance of the firm.
2. It is very costly for supervisors to assess accurately the performance of individual workers, including the use of their specialized knowledge.

In firms characterized by these two conditions, workers have the potential to engage in opportunistic behavior that enhances observers' estimates of their productivity, but in fact does not contribute to the firms' goals. In these situations, it is important to minimize the incentives for workers to engage in such behavior. A contractual system in which seniority plays a central role in determining the compensation and job security of workers contributes to this objective.

An oft cited example (e.g., Thurow, 1976) of a situation in which there is potential for opportunistic behavior is the operation of machine tools. Experienced workers acquire knowledge of the idiosyncracies of particular machine tools. Efficient operation dictates that new workers acquire this knowledge as rapidly as possible. However, since the information is not codified, it can only be transmitted to new workers quickly if experienced workers provide informal on-the-job training. If compensation and job security depend on assessments of the performance of individual workers, experienced workers have the incentive to conceal such knowledge.

Thus, the combination of specialized knowledge that can be used strategically and high transaction costs associated with monitoring the use of this knowledge creates a situation in which contracts that base compensation and job security on seniority may elicit behavior more consonant with the firm's goals than contracts that reward assessed performance.

Employment contracts influence the efficiency of firms not only by affecting the behavior of workers while on the job, but also by affecting the mobility of workers. Firms offering employment contracts that reward observed performance will attract productive workers if the following two conditions characterize the production activity:

1. The contribution of individual workers to the firm can be assessed accurately by supervisors. (In other words, there is no potential for undetected opportunistic behavior.)

2. The relationship between the worker's actions and the worker's performance is stable.

When either of these conditions is absent, employment contracts that base compensation on assessed performance may not attract the most productive workers.

Consider the first condition. In situations in which opportunistic behavior can go undetected, performanced-based contracts create conflict between behavior that leads to high monetary compensation and behavior that promotes productivity. If the job satisfaction of productive workers depends on a sense of efficacy in doing a job well (March and Simon, 1957), such workers may react to this conflict by leaving the firm.

The second condition concerns the degree of stability in the relationship between the actions of the worker and assessed output. If the relationship is unstable (the same worker actions result in different output levels at different points in time), risk averse workers will accept performance-based contracts only if the reward structure includes risk premiums to compensate workers for assuming the risk of factors beyond their control (Hirshleifer and Riley, 1979). Firms that use performance-based contracts without risk premiums will find it difficult to attract productive workers. If the instability is very great, and consequently large risk premiums would be required, firms may find it efficient to pay workers on the basis of seniority even though such contracts do not provide incentives for high effort levels.

In summary, the efficiency of alternative contracting forms is determined by the nature of the production technology and the level of transaction costs associated with monitoring performance. In situations in which there is a stable relationship between worker actions and assessed performance, and the actions of workers can be monitored at low cost, contracts that base job security and compensation on supervisors' assessments of worker competence will be efficient (Alchian and Demsetz's example of truck loading). In situations in which the output of an individual worker can be observed and evaluated at relatively low cost, contracts that reward performance will be efficient (ironing shirts). In situations in which workers acquire information critical to the productivity of the organization as they work at their jobs, and the use of this information cannot be monitored without high costs, employment contracts in which job security and compensation are heavily influenced by seniority may be relatively efficient (operating machine tools that have idiosyncratic bugs).

C. CHARACTERIZING THE PRODUCTION PROCESS IN EDUCATION

This section describes attributes of public education that influence the efficiency of alternative contracting forms. The first of these attributes of public schooling is so peculiar and subtle in its effects that its significance for economic issues has not been fully recognized. It demands special attention, not only for the limited problems discussed in this essay, but in all analyses of the economics of education. The other attributes (numbered 2-5) are characteristics of the production process determining the achievement of children. These characteristics are important in applying the analysis of the previous section to the issue of efficient contracts for public school teachers.

1. Student Diversity and the Equal Access Mandate

A central fact about the public schools is that they have a mandate to educate every child who comes to the school door. This mandate is reflected in a variety of public documents, including court decisions such as Brown v. Board of Education of Topeka (1954)[1/] and Lau v. Nichols (1974),[2/] Congressional legislation such as Title I of the Elementary and Secondary Education Act of 1965 and the Education for All Handicapped Children Act of 1975, and the provisions in many state constitutions guaranteeing that all children will be provided with a "thorough and efficient" education.[3/] While these documents do not specify exactly what is to be equalized (a point discussed later in the paper), they do imply that as a minimum every child should have equal access to the resources available in public schools, including the time and attention of teachers. I call this the equal access mandate.

The students who go to public schools, and to whom the equal access mandate applies, vary enormously in backgrounds, attitudes, skills, and handicaps. As a result of these differences, students also vary in their receptivity to school and in what they learn in school. The responses of public school teachers to this diversity and to the equal access mandate provide the focus of much of the analysis that follows.

2. Effective Teaching

There is clear evidence that some teachers are more effective than other teachers in helping children to acquire cognitive skills (Hanushek, 1979). However, very little is known about the characteristics of effective teaching. Despite a great deal of research, there is very little evidence of consistent relationships between the use of particular instructional techniques and student learning (Averch et al., 1972; Pflaum et al., 1980; Rosenshine, 1976).

A compelling reason for the disappointing research results is that the effectiveness of particular instructional techniques depends critically on the characteristics of the children in the class, on the skills and personality of the teacher, and on the nature of the interaction of students and teacher. The critical characteristics of students and teachers that influence the effectiveness of particular instructional techniques may be very subtle, and consequently cannot be identified by researchers. Teachers find effective techniques through a process of trial and error and adaptation. In other words, effective teaching is characterized by an efficient search process, rather than by careful application of well specified techniques.

3. Teacher Allocation of Time

The achievement gains that children make during a school year depend not only on the effectiveness of the teacher in using instructional time, but also on the allocation of instructional time to different children (Brown and Saks, 1975; Monk, 1979; Thomas, 1977). Among the important decisions teachers make is how to divide children into instructional groups and how much time to spend with individual children, with particular groups of children, and with the class as a whole. Recent evidence indicates that children's learning is sensitive to the amount of instruction the child receives, and that the gains from individualized instruction and instruction in groups of different sizes differ (Brown and Saks, 1979; Kiesling, 1979). Thus, the impact of the teacher on children's learning depends not only on the skill of the teacher, but also on the decisions the teacher makes in allocating time to different children.

4. Teacher Knowledge of Student Capabilities

As teachers search for instructional strategies and allocations of classroom time that will be effective for their students, they acquire information about the responses of individual children to instructional time. For example, they learn which students respond quickly to additional attention and which students respond only very slowly to large allocations of time and other resources. They also learn over a period of years which families are supportive of their teaching and which families can be called upon to respond to problems regarding their children. This information is acquired by on-the-job experience, interacting with children and their families, and much of it is not accessible to supervisors.

5. Peer Effects

The impact of school on a child's learning is determined not only by the actions of the classroom teacher, but also by the attributes and actions of the other children in the class. The precise nature of these peer effects has proved almost as elusive to researchers as have the determinants of effective teaching (Rosenbaum, 1980). However, the key point for this paper is not the precise nature of the peer effects, but rather that the amount of progress students make during a school year depends not only on the actions of the classroom teacher but also on peer group influences which are to a large extent beyond the control of the teacher.

D. EMPLOYMENT CONTRACTS FOR PUBLIC SCHOOL TEACHERS

This section considers how the five factors characterizing public education influence the responses of public school teachers to different types of employment contracts. Three types of contracts are considered:

1. contracts in which the job security and compensation of teachers are determined by supervisors' evaluations of teacher actions;
2. contracts in which the job security and compensation of teachers are determined by evaluations of the academic progress students make;
3. contracts in which the job security and compensation of teachers are determined by seniority.

These contract types correspond to the three types of employment contracts described earlier in the paper; however, they are discussed in a different order.

1. Contracts That Base Teachers' Salaries and Job Security on Supervisors' Evaluations of Teachers' Actions

This type of contract is not efficient for employing teachers because, as discussed in the previous section, there is no well defined relationship between particular teaching techniques and student learning. Effective teaching requires experimentation, and observation of an unsuccessful experiment does not provide evidence of ineffective teaching. Moreoever, the costs of monitoring are high, both because extensive observation is required to gain a sense of what a teacher is attempting, and also because the monitoring process itself may disrupt the interactions among teacher and students that result in learning.

It is important to note that evaluations by able supervisors will reveal the truly incompetent teacher who has not acquired any of the skills necessary to help children to learn and who does not respond to help in acquiring these skills.[4] It is clear that efficient operation of the public schools requires the dismissal of such teachers. However, the right to dismiss such teachers is not an issue of contract form. Even collectively bargained contracts with heavy reliance on seniority provide for the dismissal of incompetent teachers after this incompetence is documented through due process. Providing incentives for supervisors to document incompetence and to request dismissal is a serious concern. However, this concern is not an issue of contract form.

2. Contracts That Base Teachers' Salaries and Job Security on Estimates of Student Learning

Employment contracts that base the compensation of workers on assessed output provide incentives for workers to behave in a manner that produces the highest assessment. If the behavioral response consists of greater effort, the response is desirable. However, evaluating teachers on the basis of student learning provides incentives for other responses as well. Such responses may jeopardize the public school commitment to equality and may also result in less efficient provision of educational services.

The potential for dysfunctional responses stems from the nature of the teaching task, which is to help all of the students in a class to learn. Evaluating teachers on the basis of student performance requires the aggregation of the learning gains of the children in each class; and aggregation requires that weights be assigned to the progress of each child. These weights matter because in any given class some children make much more progress during a school year than other children do. As a result, the performance of a teacher relative to colleagues will be sensitive to the weights attached to the achievement of individual children and to the assignment of children to teachers.

Thus, the use of performance measures to determine salary and job security requires the assignment of cardinal weights to the learning gains of different children. There is no social process that provides such a system of weights. Public pronouncements clearly indicate a concern with the treatment different children receive in the public schools. However, they do not provide clear information about what is to be equalized, and consequently they do not reveal the weights that should be assigned to the achievement of individual children.5/ Given this situation, any set of weights used in evaluating teachers must be considered arbitrary, and consequently, it is important to consider the impact of these weights on the distribution of student achievement.

We now turn to a consideration of the responses from teachers that performance based employment contracts may elicit. The first response is lobbying for students who respond well to school instruction. Teachers who have been at a school for a number of years are in the best position to do this since their experiences in previous years provide them with information about the skills and attitudes of children in particular families. If senior teachers recruit the students that learn the most in school, then other teachers are left with children who are more difficult to teach. These teachers would be placed at a significant disadvantage in comparative evaluations, and consequently the integrity of the evaluation process is undermined.

It is possible to overcome this response by randomly assigning students to teachers. However, this eliminates any efficiency gains that result from matching children with particular needs to teachers with particular skills.

A second response that is more difficult to prevent is that teachers may allocate time within the classroom disproportionately to those students for whom additional instructional time results in the greatest increase in weighted achievement. Attempting to constrain teachers' allocation of time involves extremely high monitoring costs. Moreover, the trial and error process used to find effective teaching methods requires extensive experimentation with time use.[6]

Would teachers respond to the evaluation system by altering the amount of instruction they give to different children? The limited information that is available suggests that this may occur. However, it is possible that teachers may not respond at all to the imposition of performance-based contracts. But, if there is no response, nothing has been gained. It seems implausible that teachers would respond only by working harder and not by strategically using their knowledge of individual students' capabilities to allocate instructional time so as to maximize their performance rating.

Consequently, one must take seriously the possibility that evaluating teachers on the basis of the academic performance of their students would induce teachers to devote large amounts of time to some children in the class and very small amounts of time to other children. Which children would be neglected? This would depend on the weights used in the evaluation system and on the distribution of learning abilities of the children in the class. If the evaluation system weighted the achievement gains of all children equally and if all children had the same response to instructional time, then a system of equal weights would result in equal time allocation. The equal weights assumption is plausible; in fact, it is implicit in the most commonly used measure of performance, the average achievement gain of the children in the class. However, the assumption of a common response to instructional time is not plausible. There is ample evidence that children learn at different rates. Moreover, it is children who come to school with disadvantages such as broken homes and low family income who most commonly respond slowly to school instruction. As a result an evaluation system based on equal weights provides teachers with the incentive to allocate small amounts of time to children who are already disadvantaged as a result of environmental circumstances external to the school. Thus, a system of rewarding teachers on the basis of the academic performance of their students may undermine the elusive but very real social mandate to the public schools to provide an adequate education to all children. This is particularly troubling because the children most likely to be neglected tend to live in families that find it extremely costly to respond to neglect either by voicing dissatisfaction effectively or by leaving the public schools.

It is probable that many teachers will not respond to performance-based contracts by neglecting disadvantaged children. Many teachers enter the profession with a strong commitment to help disadvantaged children even though the response to such help is often small and slow in coming. Changes in the reward structure may not induce such teachers to change their teaching behavior. However, the experience of conflict between behavior that leads to high evaluations and behavior that reflects commitment to troubled children is highly frustrating and may lead such teachers to leave the profession. Resignations of teachers with particular concern for disadvantaged children would seriously impair the ability of the public schools to help such children.

Performance-based contracts could also increase the cost of employing teachers of a given level of effectiveness. The reason is that the achievement gains of the children in any given class depend only in part on the skills and effort level of the teacher. They also depend on factors beyond the teacher's control such as home and peer group influences. As a result even very talented teachers have years in which their students make much less progress than in other years (Averch et al., 1972, pp. 57-58; Begle and Geeslin, 1972, p. 143; Jackson, 1968, p. 125). The external influences are extremely subtle and it is very costly if not impossible for supervisors to assess accurately the extent to which the achievement gains of students are determined by peer and home factors. If teacher salaries are tightly tied to the achievement gains of students, then salaries of teachers would vary from year to year. Assuming teachers are risk averse, they would require significant salary premiums to compensate for bearing the risk of variation in student performance beyond their control.

I have suggested that as a result of particular characteristics of the production process in education, performance-based contracts may induce responses that jeopardize the equality of education provided to children in public schools and may reduce the effectiveness of the educational system. Is there evidence to support these arguments? The evidence is sparse. However, the limited information that is available suggests that the concerns expressed above are important ones.

The first piece of evidence comes from the federally funded performance contracting experiments in the early 1970's. Performance contracting in education is an attempt to foster productivity by offering financial incentives for success in helping children to acquire specific skills. In the experiments, private firms signed contracts under which they received a fixed payment for each child whose reading skills during a school year increased by at least a grade level as measured on a standardized test. Evaluations of the experiments revealed that at one site firms responded to the incentives provided in the contract by allocating more time to children of average ability than to high ability or low ability children. High achievers were neglected because they would increase their reading skill by one grade level without a significant amount of in-school instruction. Low achievers were neglected because they were unlikely to achieve the grade level increase in skills even with a great deal of instruction (Gramlich and Koshel, 1975, pp. 55-56). The experiment ended before there was time to learn how parents of neglected children would respond or whether alteration of the compensation algorithm was politically possible and whether it would have elicited a different allocation of resources. However, the experiments did show that firms willing to supply educational services on a profit making basis can be expected to respond to the incentives provided in the contractual agreements.

The second piece of evidence comes from a study by Philip Jackson (1968) in which he interviewed fifty teachers considered by their supervisors to be extremely effective. Among the issues Jackson discussed with these teachers was their attitude toward performance-based contracts (merit pay). These teachers, who presumably would gain additional income under such a system, indicated strong resistance. Several teachers indicated that they would not work under such a contractual system (p. 132). The reasons included the statement that performance depended greatly on factors beyond their control such as the mix of children in the class. Moreover, such a system would inhibit teamwork and creative responses to the needs of individual children.[7] Many teachers indicated that a system of compensation that encouraged opportunistic behavior reduced their job satisfaction even if it did result in additional income.[8]

The third piece of evidence concerns the results of performance-based layoffs. Several Massachusetts school districts responded to declining school enrollments by laying off teachers on the basis of evaluations of their performance. Interviews with teachers and supervisors in these districts (Johnson, 1980) indicated that this system caused such a debilitating reduction in morale and productivity that the districts discontinued the policy after only a short time.

3. Contracts That Base Teachers' Salaries and Job Security on Seniority

Contracts that base the salaries and job security of teachers on seniority provide no financial incentives for outstanding performance. For this reason such contracts are often criticized by analysts concerned with productivity. The emphasis in this essay is that in evaluating seniority rules it is necessary to compare the responses they evoke with the responses that alternative incentive structures evoke. Given the characteristics of public education, including the commitment to teach all children, the nature of the teaching task, and the lack of low cost alternatives to the neighborhood school for parents of disadvantaged children, the responses elicited by seniority rules may be less detrimental to the performance of public education than the responses elicited by contracts that reward assessed productivity.[9]

In light of the somewhat counterintuitive nature of this argument, it may be useful to state once again the reasons performance-based contracts are not effective in promoting the goals of public education. As public education is currently organized, disadvantaged children are heavily dependent on the professional dedication of teachers for the extensive help they need. There is clearly wide variation in the extent to which teachers provide such help under the current incentive system. However, a reward system that provides incentives to maximize the average achievement of students may lead many teachers to devote less time to disadvantaged children and may induce teachers particularly concerned with disadvantaged children to leave the public schools.

Given the arguments suggested above, it is instructive to ask whether seniority rules play less of a role in the personnel policies of private schools, and if so, why this is the case. There are two parts to the answer. First, while there is enormous variation in the personnel policies of private schools, in most schools seniority does play a role in determining compensation. For example, many private schools use a salary schedule that specifies a lower and an upper bound for all teachers with a given level of seniority. Both the upper and lower bounds increase with seniority. It is true, however, that there is often more flexibility in the salary schedules of teachers in private schools than is the case in public schools. This leads to the second part of the answer.

Two important ways in which private schools differ from public schools are that private schools select their students and parents select the schools for their children. Selection by the school reduces the variation in the abilities and attitudes of the children any given teacher is asked to work with. Selection by the parents means that parents who feel their child is neglected will withdraw the child. These selection procedures allow supervisors in private schools to apply the evaluation standard of whether teachers are successful in teaching a relatively homogeneous group of children who want to be at that school. This is not an appropriate standard for evaluating public school teachers.

E. INCENTIVES FOR EFFECTIVE TEACHING WITHIN A SENIORITY-BASED SYSTEM

A critic of seniority rules might make the following argument: While you have shown that performance-based contracts for teachers elicit dysfunctional responses, you have not demonstrated that incentives for effective teaching can be provided within the context of seniority-based employment contracts. Without such incentives, seniority-based contracts may not be superior to contracts based on performance assessments.

A comprehensive response to this argument is beyond the scope of this paper and, in fact, is beyond the scope of available evidence. This is the reason that this essay, while it identifies and illuminates many problems with alternatives to seniority-based contracts, is not intended to be a definitive defense of seniority-based contracts. With this caveat in mind, it is appropriate to respond to the hypothetical critic of seniority rules with the following two points.

First, within the general context of a seniority-based system, there are ways to attract, retain, and motivate effective teachers. For example, some states provide small grants on a competitive basis to teachers interested in pursuing special teaching projects. A recent study (McDonnell and McLaughlin, 1980) reports that these grants have been very effective in motivating teachers.[10] A second example is the creation of positions of "master teachers", to which exceptional teachers can be promoted.[11]

The second point is that success in developing and implementing methods to motivate effective teaching depend critically on the quality of relations between teachers and school district officials. The next part of the paper focuses on these relations during the last ten years.

II. SENIORITY RULES IN A REGIME OF COLLECTIVE BARGAINING AND DECLINING DEMAND FOR TEACHERS

In the last ten years, public education in the United States has been enormously influenced by two logically unconnected, but coincident phenomena. The first is the decline in student enrollments. The second is collective bargaining and the increased influence of teachers' unions. These developments have had a radical impact on relations between teachers and administrators in general, and in particular, on the interpretation and administration of seniority rules. This part of the paper focuses on the role of seniority rules in influencing educational productivity in a regime characterized by collective bargaining and a declining demand for teachers, precipitated by student enrollment declines. I begin with a brief description of the magnitude of the enrollment declines.

A. DECLINING ENROLLMENTS

Seniority rules governing the job security of teachers have impinged on the allocation of teaching resources in recent years, primarily as a result of the unprecedented decline in student enrollments in the nation's public schools. From a peak of 32.6 million in 1970, the number of children attending public elementary schools in the United States declined to 29.4 million by 1977. That this trend will continue for some time is suggested by the fact that the number of children in the first grade of public schools in the United States has decreased from 3.8 million in 1970 to 3.3 million in 1977 (Digest of Education Statistics, 1979).

The effects of enrollment declines on the demand for teachers have been particularly great because of the fiscal crises that hit many cities during the 1970's. As a result of these crises, many school districts could not cushion the impact of declining enrollments by reducing class size markedly. Instead administrators were told to contribute to the budget cutting effort by reallocating teachers and by reducing the number of teachers employed by the district.

B. THE IMPACT OF SENIORITY RULES ON THE PROCESS OF ADJUSTMENT TO DECLINING STUDENT ENROLLMENTS

Seniority rules place severe constraints on the process by which school districts adjust to declining student enrollments. The reason is that the rules determine to a large extent the pattern of transfers and layoffs that will result from a reduction in the size of the teaching force (Murnane, 1981). For example, when declining enrollments dictate the reduction in the number of teachers employed in a particular school, seniority rules determine which teacher must leave the school, and what the options of that teacher are. Most contracts state that if an opening for which the displaced teacher is qualified is not available in another school in the district, the teacher may displace ("bump") a teacher with less seniority from his or her position.

The operation of these rules threatens the efficiency of school district operations for several reasons. First, staffing patterns are often disrupted, resulting in the breakup of teams of teachers that have learned to work together effectively over a number of years. Second, the operation of these rules results in the layoff of many young teachers. Many administrators feel that this is particularly costly to the school system because as a result of the current excess supply of teachers, administrators have been able to upgrade the quality of their teaching staff by being highly selective in choosing among the many applicants for positions. In addition, young teachers tend to be more responsive to innovations and therefore the aging of the teaching population may make it particularly difficult to develop and implement new ideas (Berman and McLaughlin, 1977, p. 136).

A third problem concerns the budgetary impact of layoffs based on seniority. Since salaries are determined by seniority in most districts, layoffs of the most junior teachers provide the least relief to financially strained school districts. One final problem is that seniority-based transfers and layoffs often jeopardize attempts to racially integrate the teaching staffs of individual schools and school districts.

These problems are severe. The constraints on the adjustment process imposed by seniority rules have frustrated administrators responsible for school district operations. However, alternative methods of allocating teachers, based on assessments of their performance, also elicit responses that tend to jeopardize the efficiency and equity of school district operations. Thus, the challenge facing administrators and representatives of teachers is to find methods of adjustment to declining enrollments that mitigate the adjustment costs while retaining the advantages of seniority rules described in Part I.

C. FINDING SOLUTIONS

How effective have school district administrators and leaders of teachers' unions been in negotiating, implementing and administering solutions to the personnel problems posed by declining enrollments? There is a great deal of variation across districts in the nature of the responses. However, in many districts, the adjustment process has been characterized by conflict that reduced the effectiveness of public education.

Why has it been so difficult for administrators and representatives of teachers' unions to work out satisfactory responses to the personnel problems created by declining enrollments? One reason is that the adjustment process is more constrained than is the case in other industries. Many private sector firms faced with a decline in demand for their products can alter the size of their product inventories to buffer the impact of demand changes on employment. They can also conduct marketing compaigns to increase the demand for their products. Neither of these responses is available to school districts.

A second, more important reason for the debilitating conflict that has characterized the adjustment process in many districts is the lack of expertise on the part of participants. Most school administrators were trained during a period when the critical problem facing public schools was rapid enrollment growth. Few administrators had experience with declining enrollments. Consequently, little was known about the responses that particular adjustments would elicit.

Another reason that district level administrators were unprepared for the 1970's was that prior to collective bargaining, superintendents often acted as advocates for teachers, making the case to city councils that high quality education required higher teacher salaries. Many administrators felt uncomfortable with the change from advocate for teachers to bargaining opponent. Consultation and negotiation with representatives of teachers' unions were not part of the decisionmaking process as they had learned it.[12/] Consequently, their reactions to pressure from teachers' unions sometimes included unilateral actions that violated the spirit, if not the letter, of teachers' contracts.

A third difficulty was lack of preparation time. In the first years of collective bargaining, preparing for contract negotiations with teachers' unions was only one of many duties of district administrators, a duty often added on top of other responsibilities. Few administrators had adequate time to prepare for collective bargaining.

Many of the reasons that administrators lacked expertise in finding solutions to the problems of declining enrollments in a regime in which power is shared with teachers' organizations also pertain to teachers' representatives. In the first years of collective bargaining, many union leaders retained all or part of their teaching duties, and consequently had little time to prepare for collective bargaining. Also, few union leaders had extensive experience in collective bargaining. Prior to collective bargaining teachers relied primarily on persuasion to achieve their goals. The basic strategy was to make a persuasive case for improved salaries and working conditions. Teachers expected that if the case was a compelling one, the school district would honor their request. Collective bargaining works quite differently. Although persuasion still plays a role, exchange is the central characteristic of decisionmaking under a system of collective bargaining. In the first years of collective bargaining many union representatives, lacking expertise in exchange relationships, made a persuasive statement of their demands and then adopted a take it or leave it strategy rather than the give and take strategy that characterizes successful collective bargaining.

As a result of the lack of expertise of administrators and union leaders, adjustments to declining enrollments in the early 1970's were often characterized by a lack of trust, by the absence of meaningful consultation or negotiation, and ultimately by work stoppages, court suits and other manifestations of conflict that reduced the ability of the schools to educate children.

In many communities, the expertise of school district administrators and union leaders has increased in recent years. Administrators have learned methods of allocating resources that do not violate the letter or spirit of teachers' contracts. Union leaders have become more aware that the long-run welfare of public school teachers depends not only on new benefits, but also on positive public attitudes towards public education. Both parties have learned the importance of consultation and negotiation in coping with the large number of unpredicted personnel problems that continually arise in public education and in interpreting provisions of teachers' contracts that are often ambiguous in the face of unpredicted personnel problems.[13/]

A significant example of this increased expertise concerns the methods used to lay off teachers. Most contracts state that teachers who will be laid off at the end of the school year must be notified by a particular date, such as April 1. The logic of this rule is that early notification provides time for teachers who will be laid off to search for alternative employment. The problem this rule poses for school districts is that accurate projections of student enrollments and teacher resignations are not available by the notification date. In the past many districts responded to this dilemma by sending layoff notices to a much larger number of teachers than the district ultimately expected to lay off. Teachers' unions argued that this constituted an unfair labor practice in that it meant that teachers were forced to bear the risk of uncertain enrollment and resignation patterns. In some cases the district action precipitated work stoppages and court action on the part of teachers. Another response, unpredicted by school administrators, is that many teachers did find alternative employment, forcing the district to incur the costs of screening new applicants at the end of the summer to fill vacant positions.

In recent years district administrators and union leaders in some communities have been able to negotiate changes in the notification rule that provide benefits to both the school system and teachers. A typical negotiated change is that teachers who may be laid off may voluntarily accept a delay in notification to August 1, in return for an extension of health benefits to the end of the calendar year if layoff proves ultimately necessary. This negotiated settlement provides teachers with an important fringe benefit and allows the district additional time to acquire information about the demand for teachers before making layoff decisions. In addition to the benefits to teachers and to administrative flexibility, this solution may contribute to productivity by permitting the retention of teachers who have experience in the school system.

Early retirement programs are another example of a creative solution to the personnel problems caused by declining enrollments. These programs provide financial incentives for older, high salaried teachers to retire, thereby reducing the need for involuntary transfers and layoffs of less senior, lower salaried teachers. Like the change in notification rules, successful implementation of early retirement programs requires expertise on the part of labor and management and a general atmosphere of trust.[14/] It is difficult to believe that successful implementation of early retirement programs or changes in notification rules would have been attained in the atmosphere of confrontation that characterized labor management relations in many communities in the early 1970's.

This section has pointed out ways that seniority rules increase the difficulties school district managers face in adjusting to declining enrollments. In assessing the overall impact of seniority rules on the quality of education provided to children, these difficulties must be weighed against the advantages of seniority rules described in Part I. The key point of this part of the essay is that the ultimate impact of seniority rules on the ability of the public schools to provide high quality education to all children depends critically on the expertise of district managers and union leaders responsible for negotiating, interpreting and administering these rules.

In situations in which negotiations are carried out by expert bargainers in an atmosphere not charged with confrontation and recrimination, it appears possible to find solutions to the problems posed by declining enrollments that retain the seniority rights of teachers and also retain a considerable amount of flexibility in allocating teaching resources.

SUMMARY

Seniority rules in teachers' contracts, vigorously defended by union leaders and criticized by some officials and analysts, have been the subject of a good deal of misunderstanding. This paper has attempted to clarify the significance of seniority rules by placing them in the context of the production process of schooling. This analysis suggests that seniority rules in education are not intrinsically dysfunctional; like all conceivable institutional rules they sometimes cause problems, but these problems should be viewed in the broader context of the rules' contribution, or lack of contribution, to the performance of the sector.

The first part of the essay shows that the relative efficiency of alternative contract forms depends on

1. the definition of performance,
2. the technology of the education process,
3. the nature of the choices available to parents unhappy with the education their child is receiving.

Given the equal access mandate, the technical characteristics of the education process, and the lack of options available to low income parents, contracts that base the compensation and job security of teachers on seniority may promote the goals of public education more effectively than performance-based contracts.

The second part of this essay explains that the effects of seniority rules on the difficult process of adjustment to declining enrollments are determined by the expertise of the agents who interpret and administer those rules. Neither school district officials nor teachers' union leaders were fully prepared for the challenges posed simultaneously by declining enrollments and the introduction of collective bargaining in the 1970's. This lack of preparation is a more compelling explanation for the problems that declining enrollments created for public education in the 1970's than the type of contracts used to employ teachers.

FOOTNOTES

1. Brown v. Board of Education of Topeka, 347 U.S. 483 (1954).

2. Lau v. Nichols, 414 U.S. 563 (1974).

3. This language appears in the New Jersey state constitution and played a significant role in the New Jersey school finance case, Robinson v. Cahill, 118 N.H. Super. 223 (Law Division) 1972.

4. There is limited evidence that supervisors can identify incompetent teachers. Two studies (Armor et al., 1976; Murnane, 1975) report that principals' evaluations of teachers do reflect teaching performance as measured by student test score gains. One might argue that the results of these studies weaken the argument against basing teachers' salaries on supervisors' evaluations. However, it is important to recognize that in the districts that supplied the data for these studies, evaluations did not influence the compensation or job security of teachers. Consequently, teachers had no incentive to engage in opportunistic behavior that would have reduced the quality of education provided to children and would have reduced the ability of supervisors to assess the contributions of individual teachers. One other relevant point is that in the Murnane study, the evaluations of principals were significantly related to the performance of white teachers, but were not significantly related to the performance of black teachers.

5. At first glance, it appears that the economics literature on principal-agent relationships should provide insights on the question of efficient contracts for teachers. However, the formal models are not helpful because they posit that the principal has a clear objective function. The lack of consensus on the weights to be attached to the

learning gains of individual children is evidence that the public schools do not have a clear objective function they seek to maximize.

6. Another dysfunctional response is focusing instruction on test content. This may result in high test scores. However, if such instruction diminishes students' interest in learning for its own sake, it may reduce students' ability and desire to learn on their own in future years.

Another, much discussed response is cheating on tests. The problem of cheating is not emphasized because it is only one of several dysfunctional responses. If it were the only dysfunctional response, a tighter control system might solve the problem. A key point of this paper is that controls cannot cope effectively with many of the dysfunctional responses that a merit pay system might engender, such as neglect of particular children.

7. Seymour Sarason has stressed the importance of interaction with colleagues in helping teachers to develop productive responses to the problems they face. His recent work (1977) has emphasized the role of networks in facilitating such interaction. Performance-based contracts may hinder the development of networks of support by making teachers reluctant to share ideas and materials, and more importantly, by making them reluctant to admit problems they are experiencing. In this perspective, one might argue that contracts that emphasize seniority may be relatively efficient because they provide a necessary (although surely not sufficient) condition for creative interaction among teachers. Such creative interaction may lead to productivity increases that more than offset losses in productivity caused by the lack of tie between

productivity and salary in contracts based on seniority.

8. The Jackson evidence reflects the responses of teachers who chose to work in a seniority-based reward system. One could argue that the responses of these teachers do not provide compelling evidence about the consequences of performance-based employment contracts. The reason is that such a system might attract teachers with very different preferences.

9. The lack of low cost alternatives to the neighborhood school has been discussed as one of the reasons that performance-based contracts for public school teachers may reduce the quality of education provided to disadvantaged children. The reader may infer from this that a system that provided a range of educational alternatives would result in better education. Consideration of this complex issue is beyond the scope of this paper. However, it is important to point out that any educational system that provided poor families (as well as other families) with a meaningful range of choices would be a system involving third party payments and a significant amount of regulation. (See Educational Vouchers [1970] for a discussion of alternative models.) Thus, the relevant debate is not about the relative merits of public school monopolies and free market competition. Instead, the relevant debate concerns the properties of alternative regulatory regimes.

10. These grants differ from merit pay in that the competition is voluntary and there is, at least in principle, no limit on the number of teachers in a school or school district who may receive awards. The McDonnell and McLaughlin study (p. 100) indicates that the stimulus to performance provided by the grant program came primarily from the recognition of initiative and was quite independent of the size of the grant.

11. The creation of a hierarchy of teaching positions, with promotions dependent on perceived merit, does introduce a significant performance element into the employment contract. However, such hierarchical job structures, which are common in the private sector, differ from the normal conception of performance-based contracts for teachers in that the compensation and job security of teachers in any given step of the hierarchy are independent of performance assessments.

12. To see how different the management of human resources was before the introduction of collective bargaining, see Gerwin's (1969) description of the procedures used by the city of Pittsburgh in the early 1960's to determine teachers' salaries. Gerwin describes how Pittsburgh "granted" a general salary increase when "no comparable school districts had lower B.A. starting salaries for teachers" (p. 56).

13. See Mitchell et al. (1980) for a discussion of recent changes in relationships between teachers and administrators.

14. For a description of the role of unions in promoting productivity through participation in the management of programs such as early retirement options, see Freeman and Medoff (1979).

REFERENCES

Alchian, A. & Demsetz, H. Production, information costs, and economic organization. American Economic Review, 1972, 62.

Armor, D., et al. Analysis of the school preferred reading program in selected Los Angeles minority schools. Santa Monica: The Rand Corporation, 1972.

Averch, H.A., et al. How effective is schooling? A critical review and synthesis of research findings. Santa Monica: The Rand Corporation, 1972.

Begle, E.G. & Geeslin, W.E. Teacher effectiveness in mathematics instruction, NLSMA report no. 28. Pasadena: School Mathematics Study Group, Stanford University, 1972.

Berman, P. & McLaughlin, M. Federal programs supporting educational change, volume VII. Santa Monica: Rand, 1977.

Brown, B.W. & Saks, D.H. The production and distribution of cognitive skills within schools. Journal of Political Economy, 1975, 83.

_______. The allocation and use of instructional time in reading and mathematics: some theory and evidence. Mimeo, 1979.

Digest of Education Statistics 1979 (NCES, U.S. Department of HEW, 1979).

Education vouchers. A report on financing elementary education by grants to parents. Cambridge: Center for the Study of Public Policy, 1970.

Freeman, R.B. & Medoff, J.L. The two faces of unionism. The Public Interest, 1979, 57.

Gerwin, D. Budgeting public funds: The decision process in an urban school district. Madison: The University of Wisconsin Press, 1969.

Gramlich, E. & Koshel, P. Educational performance contracting.

Washington: Brookings, 1975.

Hanushek, E.A. Conceptual and empirical issues in the estimation of educational production functions. Journal of Human Resources, 1979, summer.

Hirshleifer, J. & Riley, J.G. The analytics of uncertainty and information--an expository survey. Journal of Economic Literature, 1979, 17.

Jackson, P.W. Life in classrooms. New York: Holt, Rinehart and Winston, 1968.

Johnson, S.M. Performance-based staff layoffs in the public schools: Implementation and outcomes. Harvard Educational Review, 1980, 50.

Kiesling, H.J. Educator objectives and school production. Mimeo, 1979.

March, J.G. & Simon, H.A. Organizations. New York: Wiley, 1958.

McDonnell, L.M. & McLaughlin, M.W. Program consolidation and the state role in ESEA Titla IV. Santa Monica: Rand, 1980.

Mitchell, D.E. et al. The impact of collective bargaining on school management and policy. Claremont: Claremont Graduate School, 1980.

Monk, D.H. Economic aspects of the allocation of resources within classrooms. Mimeo, 1979.

Murnane, R.J. The impact of school resources on the learning of inner city children. Cambridge, Mass.: Ballinger Publishing Co., 1975.

Murnane, R.J. Teacher mobility revisited. Journal of Human Resources, 1981, winter.

Pflaum, S. et al. Reading instruction: A quantitative analysis. Educational Researcher, 1980, 9.

Rosenshine, B. Recent research on teaching behaviors and student achievement. *Journal of Teacher Education*, 1976, *27*.

Sarason, S.B. et al. Human services and resource networks: Rationale, possibilities and public policy. San Francisco: Jossey-Bass, 1977.

Thomas, J.A. Resource allocation in classrooms. Final Report to the National Institute of Education, 1977.

Thurow, L. *Generating inequality: Mechanisms of distribution in the U.S. economy*. New York: Basic Books, 1976.

Williamson, O.E. *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press, 1975.

Evidence, Analysis, and Unanswered Questions:
Coleman's New Study, Public and Private Schools

by

Richard J. Murnane

Institution for Social and Policy Studies
Yale University

April, 1982

This paper is based on research supported by Grant NIE-G-79-0084 from the National Institute of Education.

Public and Private Schools is an important document that increases our knowledge of American secondary school education today. In particular, the document provides new information about private schools--the schools attended by ten percent of American high school students. Given current policy discussions about public support for private schools, it is important to learn as much as possible about these schools. The new publication by James Coleman and his colleagues contributes to this objective by providing significant information about the size and geographical distribution of different types of private schools, their curricula, and about the demographic characteristics and skills of their student bodies.

As a social scientist with an interest in the ways that the operation of markets contributes to the efficient use of resources in our society, I am both intrigued with and sympathetic to the possibility that private education and the choices it makes available may contribute to the improvement of American education. For this reason I find encouraging much of the evidence presented in Public and Private Schools. But, also as a social scientist, I am disturbed by this document because it attempts to answer questions that cannot be answered with the data available to the authors--or, for that matter, with any existing data. These questions concern the relative quality of the education provided by public and private schools and the impact of tuition tax credits or vouchers on the educational opportunities available to American children of different backgrounds.

In both its strengths and weaknesses, Public and Private Schools is reminiscent of James Coleman's earlier quantitative analysis of American education, Equality of Educational Opportunity, published in 1966. That document (commonly referred to as the Coleman Report) provided a great deal of new important descriptive information about American education. For example, the 1966 report demonstrated that

— racial segregation was not confined to the South; most children attending school in the northern part of the country also attended racially segregated schools;

— minority group children had lower average academic achievement than white children did;

— the differences in the achievement of children from different backgrounds could not be explained by observable differences in physical school resources; and

— family backgrounds were highly correlated with children's achievement.

Yet it was not the wealth of significant new descriptive information about American education that gave the Coleman Report its controversial reputation. Instead, the Report came to be associated with the conclusions that schools don't matter much, and that busing to achieve racial balance is the most effective way to improve the education of minority group children. Neither of these conclusions is justified by the analysis—the data base was simply not appropriate for analyzing the impact of school resources on individual children,[1] and the data contained no examples of busing to achieve racial balance. It took several years of reanalysis and discussion, however, before it became clear what Equality of Educational Opportunity could and could not tell us. In the interim, the document figured prominently and inappropriately in the acrimonious and painful discussions about social policy towards education.

I fear that Public and Private Schools may also play an inappropriate role in public policy discussions about education. The present policy debate concerns public support for families who choose private education for their children; this support would take the form of vouchers or tuition tax credits. It is possible, indeed likely, that this document will be presented as evidence that public support for private education will unequivocally improve the education offered to children from widely varying backgrounds. But the study does not present evidence on this point; and it does threaten to divert attention from the many critical and difficult questions related to the introduction of such aid. The purpose of this review is to illuminate the contribution that Public and Private Schools makes to our understanding of private schools and to clarify why the report does not provide reliable evidence concerning the consequences for American education of changes in public policy toward private schools.

I. WHAT PUBLIC AND PRIVATE SCHOOLS DOES TELL US

This document reports the results of detailed analyses of a new large data base describing American high school students and the schools they attend. The cross-sectional data base, which consists of information on 58,728 students attending 1,016 different high schools, is the baseline data for the National Center for Education Statistics's new longitudinal study entitled HIGH SCHOOL AND BEYOND. As a result of Coleman's extensive analyses of these data (there are 57 tables in the text), there is now information available concerning a number of questions that previously could be addressed only with fragmentary data and speculation. Examples of information provided by this document include:

1. There is wide variation across states in the percentage of high school students attending private schools. The percentages range from a low of 1.5 percent in Wyoming to a high of 17 percent in Connecticut.
2. Contrary to the results of previous work (e.g., Erickson et al., 1978), Catholic schools are concentrated in suburban communities, not urban areas.
3. The percentage of black students in Catholic schools is a little under half that in the public schools; the percentage in non-Catholic private schools is about a fourth that in the public schools.
4. Those blacks who do attend private schools are less likely to be segregated in all black or nearly all black schools than is the case for blacks attending public schools.
5. Private schools provide primarily academic programs and have few vocational or technical courses. This is true for both Catholic schools and other private schools.
6. Few American high school students receive advanced foreign language training. Only 20 percent of the students in non-Catholic private schools take a third year language course and the percentages are lower in Catholic and public schools.
7. On average, students in public schools have lower attendance rates, more behavior problems and lower achievement than children in private schools. These differences are not wholly traceable to observable differences in the family backgrounds of the students surveyed.

These are only a sampling of the many findings reported in Public and Private Schools that contribute to our understanding of American secondary school education, and especially to our understanding of the roles played today by private schools.

II. WHY PUBLIC AND PRIVATE SCHOOLS DOES NOT TELL US ABOUT THE RELATIVE QUALITY OF PUBLIC AND PRIVATE SCHOOLS: THE EFFECT OF STUDENT SELECTION

As indicated above, Public and Private Schools provides evidence that students who attend private schools have higher achievement test scores on average than children who attend public schools. Obviously, one must ask: Is this a result of differences in the quality of education provided by the schools or is it the result of differences in the skills, motivation and parental support possessed by the children who attend these schools?

We know that at least part of the difference in the average achievement of children in public and private schools stems not from differences in school quality, but from differences in the family backgrounds of children. (Coleman's is the latest of a number of studies reporting this finding.) These differences arise from the way that children are assigned to schools. If students were assigned to schools on the basis of a coin flip or a throw of a die, the average characteristics of children in public and private schools would be the same. However, in our society, school choice is determined by the decisions of parents. Family income is one important determinant of these decisions and this can be taken into account. However, it is not the only determinant. We would expect that among families with the same income, those families that make the extra financial sacrifices to send their children to private schools are those families that value education particularly highly and tend to prepare their children especially well for school. As a result of this at-home motivation and preparation, we would expect these children to have higher achievement test scores on average than children in public schools even if the quality of education provided by the two types of schools were the same. Social scientists refer to this phenomenon as self-selection.

The critical question is whether it is possible to use statistical techniques to correct for the effects of self-selection. Coleman and his colleagues attempt to make this correction by controlling for the effects on student achievement of seventeen objective and subjective characteristics of family background. Few researchers today would accept this as sound. It is now well known that the effects of self-selection cannot be adequately controlled by the inclusion of even a large number of observed family background characteristics. (This point is made clearly in many of the articles in Evaluation Studies Review Annual, Volume 6, Sage Publications, 1980. In fact, many of the articles in the 800 page volume contain discussions of alternative methods of controlling for the effects of self-selection.)[2/]

Would the use of other methods of analysis, more sophisticated than those employed by Coleman and his colleagues, have eliminated the problem of bias due to the effects of self-selection? In other words, is it possible with alternative methods of analysis to use the available data to assess the relative quality of public and private education? Social scientists will disagree on this point. However, I believe that the answer is no. Solving the problem of selection bias requires the identification and measurement of at least one variable that fulfills two conditions:

1. this variable is known to influence a family's decision concerning whether to send their child to a public or private school and
2. this variable does not influence a child's achievement.

If such a variable were found, it would be possible to use instrumental variable methods to estimate and control selection bias (Olsen, 1980).[3/] It is this reviewer's belief, however, that all observable variables that influence choice of school also influence achievement. (The education levels of parents is one obvious example of these variables.) Consequently, it may not be possible, even using more sophisticated methods of analysis than those employed by Coleman and his colleagues, to accurately determine the extent to which differences in the achievement of children attending different types of schools are due to differences in the children, or to differences in the schools.

So far the argument of this section has been based on the accepted conventional methodology for evaluating educational programs and manpower training programs in general. A critical underlying assumption of this conventional methodology is that the programs and the selection mechanisms are analytically distinct. For researchers working in this paradigm, the challenge is to develop statistical methods to account for and separate out the effects of selection mechanisms so that the effectiveness of the programs can be accurately assessed.

I would argue that the conventional methodology is inappropriate for evaluating the relative effectiveness of public and private schools. In the case of schools selection mechanisms and educational programs are not analytically distinct. Selection is an integral part of the creation of an educational program for several reasons. First, children learn a great deal from each other. One of the most effective ways to improve the cognitive skills of children is to put them in an environment with other children who want to acquire cognitive skills and whose families support such learning. Thus, selecting the right mix of students is a powerful way to improve the effectiveness of an educational program. Second, the presence of even a few disruptive students makes it difficult to maintain the sense of order and discipline that many studies, including Public and Private Schools, have shown to be positively related to learning. Therefore, the right to dismiss disruptive students, even if exercised only rarely, is an inseparable part of the process of creating order and facilitating learning. Third, the use of selection mechanisms to create a student body that accepts discipline and values learning helps in attracting and retaining talented teachers. (The history of teacher mobility in public schools can be interpreted as a history of teachers moving away from schools that face severe discipline problems and that serve numerous unmotivated children. For example, see Becker, 1952, and Greenberg and McCall, 1974.) All of these reasons support the idea that student selection simply cannot be separated from the educational process.

One implication of this view is that it is not informative to compare the quality of education provided by public and private schools. A meaningful comparison would analyze the effectiveness of the two types of schools in carrying out the same tasks with the same tools and with the same children. As Public and Private Schools documents, private schools provide primarily college preparatory instruction to those children whom they select and whose parents select them. Public schools provide instruction, both college preparatory and vocational instruction, to all children who come to the school door, irrespective of the amount of support provided by parents. Comparisons between schools facing different tasks and able to use different tools do not increase our understanding of American education, because their assumption that selection processes are simply one particular, separable aspect of schooling leads to the untested conclusion that private schools could effectively serve a different clientele than they now serve.

At several points in the document, Coleman and his colleagues show that they are aware of the importance of selection and of the way that selection mechanisms may place extra burdens on the public schools. In fact, the text of the document ends with the following statement:

> " . . . the constraints imposed on schools in the public sector (and there is no evidence that those constraints are financial, compared with the private sector) seem to impair their functioning as educational institutions without providing the more egalitarian outcomes that are one of the goals of public schooling" (p. 233).

This statement makes clear that Coleman and his colleagues are aware that the operations of both public schools and private schools are influenced by current selection mechanisms, which tend to concentrate in the public school system children who are not motivated to learn and who do not want to be in school. What is not clear from the document is how public aid for private education would influence these children. There are at least two interpretations.

The first interpretation is that as a society we have paid too much attention to these unmotivated, unruly students and have allowed them to ruin the public schools, making them places where few children can learn. Increased support for private schools is a mechanism for redressing the balance and creating a new priority based on the idea that public funds for education should be directed toward those students who want to learn. If this is the underlying motivation for aid to private schools, it should be discussed openly. The commitment to equity has been extremely costly in many respects in American public education; it has also produced many benefits. Decisions to reduce this commitment should not be made without explicit discussion and debate. Moreover, if such a decision is considered, the decision makers need to be reminded that support for private education is only one of several mechanisms for implementing such a change.

There is a second interpretation of the case for public support for private education. One might argue that many children who disrupt public schools do so because they are unhappy with the education offered to them and are frustrated by the requirement that they attend the school to which they are assigned. It is possible that many of these children might learn more and respond more positively to formal education if they and their families could choose the school that they attend. Moreover, they might respond positively to a highly disciplined environment that they chose.

This is an exciting possibility. However, to have a chance of success, such children must have real choices. In the next section I discuss some of the issues that bear on the question of whether a system of public support for private education would give real choices to all American children.

III. WHY PUBLIC AND PRIVATE SCHOOLS DOES NOT TELL US HOW TUITION TAX CREDITS OR VOUCHERS WOULD INFLUENCE THE QUALITY OF EDUCATION PROVIDED TO MINORITY GROUP STUDENTS AND STUDENTS FROM POOR FAMILIES: THE ISSUES OF SUPPLY RESPONSE AND REGULATION

Public and Private Schools provides some significant evidence about the experiences of minority group students and poor students in private schools. But how significant? The document shows that most black children in private schools attend integrated schools. It also shows that Hispanic children from higher income families are more likely to attend private schools than are Hispanic children from poor families. From these results the authors infer that public aid to parents who choose private schools for their children would increase the educational opportunities available to poor and minority group children.

These inferences about the effects of public aid are not sound for two reasons. First, Public and Private Schools provides no information about how private schools (existing ones and new entrants) would respond to the increased demand for private education that would result from tuition tax credits or vouchers. Second, when public funds have been made available for the private provision of social services in the past, this has led to significant public regulation of the private providers (for example, nursing homes). There is good reason to believe that public support for private education would follow this same course. We do not know how regulation would influence the educational opportunities of children from different backgrounds or the quality of education provided by private schools. Let us consider these points in turn.

The effects of tuition tax credits or vouchers on the educational opportunities available to poor and minority group children will depend on the unknown answers to these questions about what economists call the "supply response" of private schools to the increase in demand:

1. How would tuition charges in private schools change in response to tuition tax credits or vouchers? The vast majority of private schools are nonprofit institutions and we know very little about the pricing policies of such institutions. For example, we do not know whether the subsidies that many Catholic dioceses provide to urban Catholic schools would be reduced if public aid were provided to the parents who send their children to these schools.
2. Would the number of places in private schools expand to meet the increased demand? Would expansion take place through increases in the student bodies of existing schools or would new schools open? Would the new schools provide education of the same quality as existing private schools? We know very little about the responses of nonprofit institutions to increased demand. Particularly in the case of Catholic schools, which face an acute shortage of priests, nuns, and

brothers, it is not clear whether these schools would expand to meet increased demand, or even if it is possible for these schools to increase supply without raising tuition substantially to cover the higher cost of the salaries of lay teachers.

3. If the private schools that parents perceive as providing the best education did not expand sufficiently to meet the increased demand, how would these schools decide whom to educate and whom to exclude? Once again, we do not know the answer to this question, and the effect of public support for private schools on the quality of education provided to poor and minority group children will depend directly on the answer.

These questions about the "supply response" of private schools to the introduction of tuition tax credits or vouchers are not addressed by the Coleman study.

As I argued above, public support for the provision of social services by private providers in other sectors has led to significant regulation of the behavior of these providers. Reasons for the regulation include concern about access to these services by the poor, concern about the quality of care paid for with public funds, and concern that the suppliers may defraud the government. (Such fraud has often taken the form of collusion between the supplier and consumer; these agents request payment from the government for services not actually performed and then divide up the public payment.)

It seems extremely likely that public aid for private education would also be followed by public regulation. We do not know what form these regulations would take, how regulations would affect the educational options of poor families, or how regulations would affect the quality of education provided by private schools. It does seem clear, however, that the design of regulations poses a major public policy challenge. On the one hand, a lack of regulation of the pricing and selection policies of private schools may result in the exclusion of poor and minority group children from certain schools. On the other hand, significant public regulation may result in a reduction in the quality of education provided by private schools by hindering the use of the powerful tool of selection. In fact, the critical public policy question that advocates of tuition tax credits or vouchers must address is whether it is possible to design a system of public support for private education that retains the American commitment to equality of opportunity and at the same time leaves private schools with the flexibility to function effectively.[4]

The discussion in this section has raised issues that are not presented in Public and Private Schools. However, it is important to consider the issues of supply response and regulation because they are central to the determination of how tuition tax credits or vouchers for private education would affect the quality of education provided to American children. Until we learn how private schools would respond to an increased demand for their services and to the regulations that would accompany (even indirect) public support, we cannot predict how such support would affect the quality of education provided to children from different backgrounds.

SUMMARY

This review is written both to illuminate the important contribution that Public and Private Schools makes and to clarify why it does not provide reliable evidence about the consequences of significant changes in public policy toward private schools. It is time for analysts and policymakers to go beyond comparisons of public and private schools and to begin to focus on the issues of selection, supply response and regulation. It is these issues that determine how public aid to parents who choose to send their children to private schools would affect education in America.

Footnotes

1. See Murnane (1975) for an explanation of the limitations of the Equality of Educational Opportunity survey.

2. For a particularly lucid explanation of the problems created by self-selection, see the article by Barnow et al. entitled "Issues in the Analysis of Selectivity Bias" in Stromsdorfer and Farkas (1980).

   It is important to note that while it is known that the Coleman methodology does not produce consistent estimates of the relative effectiveness of school programs, it is not known whether the use of this methodology undercompensates or overcompensates for differences in the skills and motivation of children attending public and private schools.

3. In principle there is a method for controlling for selection bias that does not require fulfillment of the conditions described in the text. However, as Olsen (1980) has explained, when this alternative method (the Mills' ratio method) is applied in situations in which the conditions described in the text are not fulfilled, the estimated coefficients become unstable and inference is very difficult.

4. Many of these issues are discussed in Education Vouchers (1970).

References

Becker, H. The career of the Chicago public school teacher. American Journal of Sociology, March 1952, 57.

Coleman, J.S. et al. Equality of Educational Opportunity. Washington, D.C.: Office of Education, National Center for Education Statistics, 1966.

Education Vouchers. A Report on Financing Elementary Education by Grants to Parents. Cambridge: Center for the Study of Public Policy, 1970.

Erickson, D.A., Nault, R.L. & Cooper, B. Recent enrollment trends in U.S. Nonpublic schools. In S. Abramowitz and S. Rosenfeld (Eds.) Declining Enrollments: The Challenge of the Coming Decade. Washington, D.C.: National Institute of Education, 1978.

Greenberg, D. & McCall, J. Teacher mobility and allocation. Journal of Human Resources, 1974, 9, 480-502.

Murnane, R.J. The Impact of School Resources on the Learning of Inner City Children. Cambridge: Ballinger, 1975.

Olsen, R.J. A least squares correction for selectivity bias. Econometrica, 1980, 48, 1815-1820.

Stromsdorfer, E.W. & Farkas, G. (Eds.) Evaluation Studies Review Annual. Volume 5. Beverly Hills: Sage, 1980.

The Uncertain Consequences of Tuition Tax Credits:
An Analysis of Student Achievement and Economic Incentives

by

Richard J. Murnane

Institution for Social and Policy Studies
Yale University

June 18, 1982

This paper is based on research supported by the Institute for Research on Educational Finance and Governance, Stanford University, by the Program on Nonprofit Organizations, Yale University, and by Grant Number NIE-G-79-0084 from the National Institute of Education. The author would like to thank Marc Chupka, George Cohen, and Stuart Newstead for first-rate research assistance, Randall Olsen and Jon Peck for statistical advice, and Paul DiMaggio, Michelle Hirschoff, Estelle James, Richard Nelson, Edward Pauly, Susan Rose-Ackerman, and John Simon for helpful comments on earlier drafts.

ABSTRACT

The debate over tuition tax credits has neglected three factors that would critically determine the consequences of such a policy for the distributions of educational achievement and education costs. These factors are student body composition effects, the interdependence of the public and private school sectors, and the incentive effects of third party payments. An analysis of tuition tax credits that does pay attention to these factors reveals that the consequences of tuition tax credits would be extremely sensitive to the details of the regulations defining the system. Moreover, any system of tuition tax credits would be characterized by powerful tensions among the public policy goals of access, minimum quality standards, and the prevention of fraud and cost increases.

I. INTRODUCTION

The issue of tuition tax credits for parents who send their children to private elementary and secondary schools has been hotly debated in recent years. However, the quality of the debate has not matched the intensity of the rhetoric. A critical reason for the poor quality of much of the discussion of tuition tax credits—and of public policy toward private schools, in general—is the lack of attention paid to three factors:

1. the role of student body composition in fostering school effectiveness,
2. the nature of the interdependence of the public and private school sectors,
3. the incentive effects created by public support (direct or indirect) for the private provision of education.

The goal of this paper is to provide an analysis of tuition tax credits that focuses on these three factors. This analysis neither attacks nor defends tuition tax credits. Instead, it develops the following three themes:

1. Observed differences between public and private schools today do not provide reliable predictions of the consequences of policies such as tuition tax credits that provide incentives for greater use of private schools.
2. The impact of tuition tax credits on the distribution of educational achievement and on the distribution of education costs would be extremely sensitive to the details of the regulations defining the system.
3. Any system of tuition tax credits would be characterized by powerful

tensions among the public policy goals of access, minimum quality standards, and the prevention of fraud and cost increases.

Section II describes the nature of the evidence on the role of student body composition in fostering school effectiveness and explains how differences in the regulations facing public and private schools lead to differences in the average composition of public and private school student bodies. Section III demonstrates that these differences in student body characteristics play a central role in explaining the differences between the achievement of students in public and private schools that were reported in Coleman, Hoffer and Kilgore's recent highly publicized study (1981a, 1981b). Section IV analyzes the consequences of tuition tax credits, emphasizing the ideas developed in Sections II and III and also the incentive effects created by public support for the private provision of education.

II. THE IMPORTANCE OF STUDENT BODY COMPOSITION

A. A Determinant of School Effectiveness

Over the last fifteen years a great deal of quantitative research has examined the determinants of school effectiveness. Among the findings of this research are that teachers matter, school programs matter (especially when defined broadly to include homework and discipline practices), and the composition of the student body matters.[1] This essay focuses on the third of these factors.

The role of student body composition first gained prominence with the publication in 1966 by Coleman and colleagues of Equality of Educational Opportunity, which emphasized the positive relationship between the average socioeconomic status of the students in a school and the academic achievement of individual students. Subsequent studies (Hanushek, 1972; Henderson et al., 1978; Summers and Wolfe, 1977; Winkler, 1975) have either replicated Coleman et al.'s finding, or have found the achievement of individual students to be related to other characteristics of the student body—such as racial composition, and the average achievement level of the students in the class or school.

Experts differ in their explanations of why average characteristics of student bodies are systematically related to the achievement of individual students, and attempts to sort out alternative explanations have not been very fruitful. However, for the purpose of this essay, the important fact is not why the composition of the student body matters, but rather that parents and school officials know that it does matter, and this knowledge influences their actions. In particular, student body composition effects create incentives for parents, in choosing schools for their children, to pay attention to who the classmates will be, and incentives for school officials to pay attention to the attributes of the students they admit. The actions of parents and school officials, taken in the context of a number of institutional constraints, result in significant sorting of students among schools. It is important to understand the nature of this sorting because it creates differences in the characteristics of the student bodies of different schools, and since student body compoisition influences school effectiveness, it also influences the achievement of students attending different schools.

B. Sorting Mechanisms Within the Public and Private School Sectors

The problem facing parents is to find a school for their child that has a student body that will enhance their child's education—as they define it. This problem is made complex by the difficulty of collecting reliable information about students in particular schools. Given this information scarcity, many parents, in making school choices, pay attention to variables such as the socioeconomic status of students, which is quite easily observed and which is correlated, albeit imperfectly, with less easily observed variables--such as the skills and attitudes of students--that help schools to foster cognitive achievement.

Of course, not all families succeed in placing their child in the type of school they desire. There are three types of sorting mechanisms at work in both the public and private sectors that determine which families succeed in placing their children in effective schools--schools that are usually, but not always, characterized by student bodies with relatively high socioeconomic status. These sorting mechanisms are self-selection by families, admission policies of individual schools, and dismissal policies of schools.

Self-selection in the public sector takes place primarily through residential location, since in the majority of public school districts in this country, the school a child attends is determined by a family's residential location. There is compelling evidence that families pay premiums for housing in school districts with reputations for good schools (Edel and Sklar, 1974; King, 1977; Oates, 1969, 1973; Pollakowski, 1973; Reinhard, 1981).[2/] There is also evidence that families pay premiums to live in neighborhoods within the same school district that are served by schools in which the average achievement of the students is high (Grether and Mieszkowski, 1974).

Selective admission policies are used by some public schools. In addition to well known examples such as the Bronx High School of Science, there are now many schools in low income areas that employ achievement criteria for admission (Fuerst, 1981).

Dismissal of disruptive students is also a method of sorting that is used in the public sector. All public school districts have procedures that school administrators can use to suspend, and, if necessary, to expel students who consistently violate school rules.[3/]

In addition to these mechanisms that sort students into different schools, there are also mechanisms, often referred to as ability grouping and tracking, that sort students into different groups within public schools (Rosenbaum, 1976, 1980). All of these mechanisms influence the distribution of student achievement because they influence the nature of the peer group with which individual students interact.[4/]

The substantial extent to which sorting takes place within the public sector is indicated by the distribution of the socioeconomic status (SES) of high school seniors attending public high schools in 1980. Based on a sample of data from the High School and Beyond (HSB) study (which is described below), 25 percent of the variation in SES among public high school seniors consists of differences among school averages.[5/] In other words, sorting of students by socioeconomic status among public schools results in significant differences in the average SES of the student bodies of U.S. public high schools.

Many parents who do not find in the public schools available to them the teachers, programs, and peer groups that they want for their children choose private schools. Parents' choices of private schools, constrained by their incomes and by the tuition, admission, and dismissal policies of individual private schools, also lead to significant sorting of students among private schools. For example, 29 percent of the variation in the socioeconomic status of seniors attending Catholic high schools in the HSB sample consists of differences in average SES between schools. The analogous percentage for seniors in private non-Catholic schools is 48 percent.

C. Sorting Between the Public and Private School Sectors: The Role of Differential Regulations

Clearly, sorting takes place not only within the public and private school sectors, but also between the two sectors. The nature of this sorting, which is critical in understanding differences between public and private schools, is influenced not only by the three types of sorting mechanisms described above, but also by two sets of regulations The first set consists of compulsory education statutes that mandate that all children up to a certain age must receive formal schooling. The second set consists of laws that guarantee the right of all students to a free education in a public school. These laws have the effect of strictly defining the total size of the student population and dictating that all students who are not sorted into the private sector must be educated in public schools. In other words these regulations affect the student populations from which student bodies in public schools and private schools are drawn.

One example of how these regulations affect the student population from which the student bodies of public and private schools are drawn concerns the treatment of students dismissed from particular schools. Schools in both the public and private sectors dismiss troublesome students (although dismissal is more difficult in public schools as a result of due process requirements). The difference is that students dismissed from a private school need not be accepted by another private school.[6] These students, however, like students dismissed from a public school, are entitled to an education in a public school.

In other words, the distribution of students between the public and private sectors and, since student body composition matters, the distribution of student achievement, are influenced by regulations that apply differentially to the public and private sectors in our two sector educational system.

III. THE ROLE OF SORTING IN EXPLAINING DIFFERENCES BETWEEN PUBLIC AND PRIVATE SCHOOLS

A. Understanding Recent Evidence: Two Experiments

In 1981, Coleman, Hoffer and Kilgore (henceforth, CHK) completed a highly publicized study of public and private schools and the students who attend these schools. Their analysis was based on the first wave of data from the High School and Beyond (HSB) project, a federally funded study of 58,728 students attending 1,016 different public, Catholic, and non-Catholic private schools. (This last group will be referred to as other private schools.) A central focus of CHK's analysis was a comparison of the relative effectiveness of public and private schools. They interpret their analysis as providing evidence concerning the following hypothetical experiment:

> Compare the achievement of the average student in public schools with the achievement that student would have if he or she attended a Catholic or other private school.

Several critics (Bryk, 1981; Fetters et al., 1981; Goldberger and Cain, 1982; Murnane, 1981a; Noell, 1982; Page and Keith, 1981, Willms,

1981) have argued that CHK's analysis strategy does not provide reliable evidence concerning the result of this hypothetical experiment. This paper, however, does not challenge CHK's methodology. Instead, it essentially replicates their analysis for a subset of the HSB data and then extends their methodology to consider a second hypothetical experiment:

> Compare the achievement of the average student in public schools with the achievement that student would have if he or she attended a Catholic or other private school and took along his or her public school classmates.

The sample drawn from the HSB database for the analysis described in this paper includes all seniors attending Catholic schools, all seniors in non-elite other private schools, and all seniors in a random sample of 123 public schools. The size of the public school subsample was chosen to make the number of public school students roughly comparable to the total number of Catholic and other private school students in the sample. Only students with complete data were included in the analysis.

Following CHK's general strategy, I estimated equation (1) for each of the three subsamples:

$$A_i = \sum_{j=1}^{10} d_j D_{i,j} + a_1 \qquad (1)$$

where

A_i = the ith high school senior's raw score (total number right) on a 47 item test of vocabulary and reading skills. (The test score was actually the sum of the scores on three subtests in the HSB student questionnaire.)

D_{ij} = the jth demographic characteristic of the ith student (the 10 background characteristics are listed in Table 1).

Table 1

Means and Standard Deviations (both weighted and unweighted) for Variables Used in Estimating Equation (1) and Equation (2)[a]

School Type	Public		Catholic		Other Private	
	Unw.	Wei.	Unw.	Wei.	Unw.	Wei.
Score on Test of Reading and Vocabulary Skills	23.72 (8.50)	24.44 (8.39)	26.41 (8.26)	27.83 (8.00)	28.50 (8.95)	29.14 (9.2)
Index of Socioeconomic Status	-0.14 (0.72)	-0.05 (0.70)	0.10 (0.72)	0.23 (0.67)	0.45 (0.74)	0.50 (0.76)
Student is Black	0.10 (0.30)	0.07 (0.27)	0.13 (0.34)	0.05 (0.24)	0.02 (0.16)	0.02 (0.15)
Student is Male	0.46 (0.50)	0.46 (0.50)	0.41 (0.49)	0.40 (0.49)	0.46 (0.50)	0.44 (0.50)
Mother wants Student to Attend College (Student's Perception)	0.65 (0.48)	0.64 (0.48)	0.83 (0.38)	0.81 (0.39)	0.77 (0.42)	0.77 (0.42)
Father wants Student to Attend College (Student's Perception)	0.57 (0.49)	0.58 (0.49)	0.75 (0.43)	0.76 (0.43)	0.74 (0.44)	0.74 (0.44)
Student is of Hispanic Background	0.14 (0.34)	0.08 (0.27)	0.20 (0.40)	0.07 (0.25)	0.01 (0.07)	0.01 (0.07)
Student has Two Parents Living at Home	0.73 (0.44)	0.75 (0.43)	0.79 (0.41)	0.84 (0.37)	0.81 (0.39)	0.79 (0.40)
Student Lives in Northeast Region of U.S.	0.12 (0.32)	0.13 (0.33)	0.35 (0.48)	0.37 (0.49)	0.16 (0.37)	0.27 (0.44)
Student Lives in North-central Region of U.S.	0.29 (0.45)	0.32 (0.47)	0.32 (0.47)	0.39 (0.49)	0.20 (0.40)	0.13 (0.34)
Student Lives in Southern Region of U.S.	0.38 (0.49)	0.34 (0.47)	0.18 (0.38)	0.13 (0.33)	0.46 (0.50)	0.39 (0.49)
Average Socioeconomic Status of Students in the School	-0.15 (0.35)	-0.07 (0.33)	0.14 (0.40)	0.26 (0.33)	0.48 (0.44)	0.53 (0.46)
Number of Students in Sample	2,464		2,026		361	
Number of Schools in Sample	123		79		21	

[a] As explained in the Appendix, equations (1) and (2) were estimated using unweighted data. However, the predicted test scores reported in Table 3 were calculated using the weighted means for the public school subsample. The weights correct for the oversampling of schools with particular characteristics, especially high percentages of minority group students. In principle, the weighted means reported in this table provide consistent estimates of the average characteristics of students in the U.S. attending schools in each of the three sectors.

(The small differences between my strategy and CHK's are described in the appendix.)

These regression results (which are displayed in columns 1, 3, and 5 of Table 2) were then used to predict test scores for a student with a given set of characteristics who attended school in each of the three school sectors. The characteristics given to the hypothetical student were the average values of the characteristics of students attending public schools. (These values are displayed in Table 1 along with the average characteristics of students attending Catholic schools and other private schools.) As was the case with CHK's estimates, the test scores predicted for the average public school student if he or she were to attend a Catholic school or other private school are higher than the average score in public schools.

To conduct the second hypothetical experiment, I estimated equation (2) for each of the three subsamples:

$$A_i = \sum_{j=1}^{10} d_j D_{i,j} + s(SESM_i) + a_2 \qquad (2)$$

where $SESM_i$ = the average socioeconomic status of the students in the ith student's high school. (This value was calculated using the information on all students in the HSB sample, both sophomores and seniors, who attended the ith student's high school.)

The estimates (which are displayed in columns 2, 4, and 6 of Table 2) were then used to predict test scores for the average public school student if that student were to attend school in each of the three sectors along with

Table 2

Estimates of Parameters of Equations (1) and (2), Estimated for Samples of Students in Public, Catholic and Other Private Schools

Dependent variable is student's score on a test of reading and vocabulary skills

(Standard errors in parentheses)

	Public		Catholic		Other Private	
	Eq.1	Eq.2	Eq.1	Eq.2	Eq.1	Eq.2
Index of Socioeconomic Status	2.32 (0.24)	1.96 (0.26)	1.96 (0.27)	1.14 (0.30)	4.15 (0.63)	2.16 (0.76)
Student is Black	-6.07 (0.53)	-5.84 (0.53)	-4.57 (0.54)	-4.06 (0.54)	-5.49 (2.72)	-4.51 (2.66)
Student is Male	0.84 (0.30)	0.85 (0.30)	1.39 (0.35)	1.16 (0.35)	0.89 (0.84)	0.61 (0.82)
Mother wants Student to attend college[a]	2.31 (0.45)	2.32 (0.45)	2.58 (0.59)	2.44 (0.58)	3.82 (1.57)	3.49 (1.53)
Father wants Students to attend college[a]	1.76 (0.45)	1.75 (0.44)	0.78 (0.54)	0.68 (0.53)	0.22 (1.54)	0.09 (1.50)
Student is of Hispanic Background	-4.23 (0.48)	-3.78 (0.50)	-3.61 (0.48)	-2.82 (0.49)	-3.35 (5.69)	-3.16 (5.54)
Student has Two parents	-0.50 (0.35)	-0.47 (0.35)	-0.00 (0.46)	0.02 (0.45)	0.99 (1.11)	1.65 (1.09)
Student Lives in North-east Region of U.S.	-0.35 (0.55)	-0.21 (0.55)	0.24 (0.55)	1.05 (0.56)	3.70 (1.46)	5.34 (1.47)
Student Lives in North central Region of U.S.	0.06 (0.45)	0.09 (0.44)	-0.29 (0.56)	-0.04 (0.56)	2.54 (1.42)	4.67 (1.46)
Student Lives in Southern Region of U.S.	-1.26 (0.41)	-0.92 (0.42)	-0.17 (0.62)	-0.80 (0.62)	-0.15 (1.21)	1.11 (1.21)
Average Socioeconomic Status of Students in the School		1.93 (0.53)		3.53 (0.58)		5.89 (1.31)
Intercept	23.19 (0.50)	23.17 (0.50)	24.26 (0.72)	23.62 (0.73)	21.45 (1.52)	18.16 (1.65)
R^2	0.23	0.24	0.14	0.15	0.25	0.29
Number of Students	2464		2026		361	
Number of Schools	123		79		21	

---

[a] As CHK (1981a) point out, this may not be an exogenous variable. Using CHK's rationale, it was included to control as completely as possible for student background.

the student's school classmates. The predicted test scores are displayed in the lower half of Table 3.

Comparison of the predicted test scores from the two hypothetical experiments suggests that all of the difference in the predicted scores between other private schools and public schools and 60 percent of the difference between Catholic schools and public schools are due to student body effects.

B. Why the Results Differ

Why do the results of the two hypothetical experiments differ? In other words, why does it matter whether the average public school student takes along his or her classmates if he or she moves to a Catholic school or other private school? There are two related reasons:

1. Student body composition (as characterized by average socioeconomic status of the students in a school) is just as important in determining the effectiveness of private schools as it is in determining the effectiveness of public schools. In fact, the regression results displayed in columns 2, 4, and 6 of Table 2 suggest that student body composition has an even greater impact on the achievement of students in Catholic schools and other private schools than it does on the achievement of students in public schools.
2. As shown by the data in Table 1, the socioeconomic status of public school student bodies is lower on average than that of the student bodies of Catholic schools or other private schools. Thus, in the case of the second hypothetical experiment, in which Catholic schools and other private schools are constrained to work with the

Table 3. Predicted Achievement Scores for Students Attending Public, Catholic, and Other Private Schools: The Hypothetical Effects of School Type and Student Body Composition on an Average Public School Student[a]

	Public Schools	Catholic Schools	Other Private Schools
First hypothetical experiment: the effect of moving a public school student to a private school			
1. Predicted achievement for an average public school student attending the specified type of school	24.3[b]	26.1	25.5
2. Predicted increase in achievement when an average public student moves to a private school (Standard error of the estimate in parentheses)		+1.8 (0.3)	+1.2 (0.7)
Second hypothetical experiment: the effect of moving a public school student and his or her classmates to a private school			
3. Predicted achievement for an average public school student attending the specified type of school with his or her public school classmates	24.4	25.1	23.4
4. Predicted increase in achievement when an average public school student moves with his or her classmates to a private school (Standard error of the estimate in parentheses)		+0.7 (0.3)	-1.0 (0.8)

Notes.

a) See Appendix for a detailed description of estimation procedures.

Achievement test scores are based on a 47-item test of reading and vocabulary. The scores in this table are the predicted number of items answered correctly.

Data are for high school seniors: 2,464 from 123 public schools, 2,026 from 79 Catholic schools, and 361 from 21 other private schools.

b) As explained in the appendix, the average characteristics of public school students used in predicting the test scores reported in Table 3 are not the average characteristics of the sample used in estimating the determinants of achievement for public school students (equations (1) and (2)). Consequently, the test scores predicted for the hypothetical average public school student attending a public school in the two experiments need not equal each other or be equal to the average of the test scores for the public school student population as reported in Table 1.

student bodies presently attending the public schools, their predicted effectiveness is markedly diminished.

The predicted achievement values presented in Table 1 should be viewed as only illustrative since they are sensitive to the methods used to generate the predictions.[7] However, these estimates do point out the importance of student body composition in explaining the achievement of students attending public and private schools. Also, it is important to keep in mind that, while the first hypothetical experiment may illuminate the choice facing an individual family about where to send their children to school, the second hypothetical experiment is of greatest relevance to the policymaker concerned with the effects of changing the school attendance choices of substantial numbers of American children.

C. Limitations of the Market Model

In concluding this section, it is important to explore how student body composition effects limit the usefulness of what might be called the market model for understanding the relative performance of U.S. public and private schools. This model emphasizes the efficiency-enhancing aspects of competition among private schools and the wastefulness of government bureaucracies. Advocates of this model (Friedman, 1962; West, 1981) explain the relatively low achievement of many students in public schools as stemming from a lack of competitive pressure on public schools.

While there are important insights to be gained from the market model—particularly in formulations that emphasize dynamic elements such as innovativeness and responsiveness to changing conditions (Nelson, 1981), this model has distinct limitations for understanding public and private schools in the United States.

First, one of the properties of normal competitive markets, that consumers are free to purchase any good for which they are willing to pay the announced price, does not hold in the case of markets for educational services. Instead, families find that their educational options depend not only on their income and willingness to pay, but also on the attributes of their children. Children who lack the attributes that particular private school managers feel contribute to the skills of the other students in the school find it difficult to gain acceptance to those schools. The reason stems from the importance of the composition of the student body in determining school effectiveness. If schools charge all students the same price, schools that do not discriminate among applicants on the basis of their effects on other students will lose desirable students to schools that do discriminate. Thus, a consequence of student body composition effects is that freedom to choose has a much more restricted definition for education than for other goods.[8/]

A second limitation of the market model is that it does not take into account the interdependence of the public and private sectors that results from compulsory education statutes and laws guaranteeing the right of every child to an education in a public school. The differential application of these regulations to the two sectors has the effect of assigning a disproportionate number of difficult students to the public sector. (Imagine the "market" outcomes if government insurance regulations required Aetna to sell insurance to all customers rejected by Prudential!) These omissions limit the usefulness of the market model in predicting the consequences of changes in public policies toward private schools.

IV. ANALYZING THE CONSEQUENCES OF TUITION TAX CREDITS

In recent years there has been a growing interest in the introduction of tuition tax credits for parents who choose to send their children to private schools. Under a tuition tax credit system, parents subtract from their federal income tax bill a part of the amount that they paid in private school tuitions. Current proposals suggest credits ranging from $250 to $500 per child. This section examines issues implicitly raised by tuition tax credits. The goal is to make three related points.

First, the descriptions of public and private schools provided by studies using the HSB and other survey data do not provide a reliable basis for judging the likely outcomes of a tuition tax credit system. The reason is that differences between public and private schools today to a large extent result from the mixed two-sector nature of the U.S. educational system and from the differential application of regulatory constraints to the new sectors. The statutes and regulations defining a program of tuition tax credits would inevitably alter the regulatory environment in which private (and possibly public) schools operate. These changes in the regulatory environment might well result in significant changes in the way sorting takes place in the public and private school sectors, and consequently in significant changes in the distributions of educational achievement and per pupil costs.

Second, the consequences of tuition tax credits for American education are extremely sensitive to the specific details of the statutes and regulations that define the system. In particular, the answers to the following crucial policy questions would depend directly on the details of the regulations:

1. Would tuition tax credits provide increased access to private schools for children from low socioeconomic status families?
2. Would it be possible in a system of tuition tax credits to assure that the credits were used only at schools that met minimum quality standards?
3. Could a program of tuition tax credits be implemented that would avoid significant cost increases and fraudulent use of tax dollars?

The third point is closely related to the second: any system of tuition tax credits would be characterized by powerful tensions among the public policy goals of access, minimum quality standards, and the prevention of fraud and cost increases. These tensions stem from the importance of student body composition in determining school effectiveness, and from the incentive effects of third party payments. The next paragraphs explore the nature of these tensions.

A. Access

Would tuition tax credits provide new high quality educational opportunities for many students, particularly for those students currently sorted out of high-achievement public and private schools? The answer depends on the extent to which the program would reduce the cost of high quality private schools to families of sorted out students, and on the responses of private schools to an increased demand by these families.

One key factor influencing the extent to which tuition tax credits would increase the demand for private education from the sorted out group is whether the plan provides net tax refunds for families that have such low income that they do not have positive income tax liabilities. Without such refunds there would be no benefits for many sorted out students.

There is no hard evidence on which to base predictions about the supply responses of private schools to an increase in demand for their services. However, it is useful to consider the potential responses of different kinds of providers.

The HSB data indicate that Catholic schools in 1980 provided education to a substantial number of minority group children and children from low income families. Would tuition tax credits result in an increase in the number of such children attending Catholic schools?

In recent years Catholic schools have experienced great financial stress, caused in part by the necessity of substituting lay teachers for increasingly scarce sisters, priests, and brothers.[9/] Many Catholic school administrators hope that tuition tax credits will ease their financial plight and allow them to keep their schools operating. Tuition tax credits for parents would improve the fiscal situation of most Catholic schools only if tuitions were raised.[10/] If tuitions were increased by the full amount of the credit, then families that cannot afford Catholic schools without the credit would still find them unaffordable with the tax credit.

Thus, the principal effect of tuition tax credits on Catholic schools may be to make them more economically viable. This may be an important consideration since the HSB data indicate that many Catholic schools do provide a good education to many children from varied backgrounds. However, to the extent that tuition tax credits lead to significant tuition increases in Catholic schools, they would not result in an expansion of the number of students from low income families served by these schools.

Would other private schools respond aggressively to an increase in the demand for private education? On the basis of economic theory we would expect the most rapid supply response to come from for-profit schools; a surprisingly large number of these schools already exist--the 1977 Census of Service Industries reports the existence of 2,237 for-profit elementary and secondary schools in the U.S. We know very little about these schools except that they are quite small (their average number of employees is 11, compared to 30 for non-profit secular schools), that they pay low salaries (their average salary in 1977 was $5,604 compared to $7,483 in non-profit secular schools), and that they are disproportionately located in the South.[11/] Theory would suggest that these schools would respond to an increase in demand by creating more places for students. However, such schools may find that the profit maximizing strategy is to specialize in serving a particular type of student--which in practice may mean students from a particular socioeconomic status. The reason is that, given parents' concern with student body composition and the difficulty of collecting information about the skills and attitudes of students, schools that accept a large number of students from low socioeconomic status families may find it difficult to attract students from more affluent families. Consequently, if schools charge all students the same price, profit maximizing schools cannot be counted on to provide low socioeconomic status families with schooling options that include schools serving students from more advantaged backgrounds.[12/]

What about non-profit private schools other than Catholic schools? The number of such non-profit schools has increased in recent years. Not much is known about the nature of their objectives and consequently about their probable responses to increased demand in general, or about their response to increased demand from low income families.[13/] (Some models suggest that they would respond to increased demand by becoming more selective in choosing among applicants (James, n.d.).)

If low SES students continue to have poor access to high quality private schools after tuition tax credits have been introduced, a number of possible regulatory issues arise: should private schools be required to admit a certain percentage of their students by lottery? should the rights of these schools to exclude disruptive students be limited? On the one hand the case for wide access to schools that receive public support (even if indirectly) seems compelling. On the other hand, as explained above, control over admission and dismissal policies are aspects of the production process of private schools that are critical in allowing them to educate students effectively. Consequently, regulations designed to assure wide access to private schools may have deleterious effects on the quality of education provided by these schools.

B. Quality

In the past, public support for the provision of social services by private providers has been accompanied by regulations designed to assure that minimum quality standards are met. The recent history of public subsidy and regulation of nursing homes in the U.S. provides an example of the pressures for such regulation, and the government response. While there is no assurance that this would take place in the case of tuition tax credits, the existence of some extremely low achieving schools among the private schools in the HSB sample (particularly among private schools not in the Catholic or the "elite" private school categories) suggests that there would be pressure for such regulations.[14/] At a minimum, pressures for making student test scores public available—so that consumers, especially government-subsidized ones, can protect themselves—seems inevitable, although likely to be resisted.

Within the public sector, regulations to assure service quality have taken the forms of certification requirements for teachers and regulations on maximum class sizes. While these regulations may have eliminated some abuses, the statistical evidence suggests little or no relationship (within observed ranges of variation) between teacher credentials or class size and student achievement (Hanushek, 1981). Moreover, these regulations have increased the cost of public education in many communities by limiting the supply of teachers and by restricting administrators' flexibility in allocating resources. Some school officials have argued that those constraints undermine their schools' ability to compete with less heavily regulated private schools.

The basic problem in regulating school quality is that it is not feasible to carry out this mission by placing controls on the outputs of schools (student skills), because these skills are influenced by many factors not under the control of the school. Controls on inputs (such as teacher certification and class size) are of limited value because the relationships to student achievement of observable and controllable factors are quite weak. As a result, regulations constraining the use of inputs in private schools could well result in significant cost increases with little positive influence on service quality.[15]

C. Fraud

A tuition tax credit plan constitutes a system of third party payments in which the third party, a unit of government, pays indirectly through tax relief for all or part of the educational services provided by a private school to a student. The incentives in a third party payment system are very different from the incentives in a simple market (in which the party receiving the services directly pays the full bill for these services). This point has been neglected by analysts who use the market metaphor to argue why tuition tax credits would unequivocally improve American education and reduce costs. In the case of third party payment systems, there are incentives for corrupt suppliers of the services to collude with consumers to defraud the third party. In the case of tuition tax credits, this collusion could take the form of a family claiming a tax credit for services not actually provided and then "purchasing" documentation certifying the child's attendance at a school in return for a portion of the tax credit.[16/]

Recent U.S. experiences with third party payment systems in the health and nursing home industries indicate that the potential for fraud is very great. In these sectors, fraud has been fought by the promulgation and enforcement of a variety of regulations: some require that private providers of services meet certification requirements (designed to assure that professional ethics rather than unconstrained profit-seeking guide administrative behavior); others require suppliers to document that they actually supplied the services for which reimbursement from the government was claimed. Analysts differ in their assessments of the effectiveness of these regulatory policies in counteracting fraud. However, it is clear that these policies have increased the cost of providing the services.

It is not possible to predict how prevalent the incidence of fraud would be if tuition tax credits for parents who send their children to private schools were introduced. It is important to point out, however, that the incidence of fraud may be related in a systematic way to the policy objective of providing access to private schools for low income families. Since not-for-profit providers are ostensibly not motivated by the search for profits, such providers may not engage in the fraud made possible by third party payments. However, these providers also have no clear incentives to expand the supply of services in the face of increased demand. For-profit providers have clear incentives to increase supply in response to an increase in demand. However, the profit motive that induces the supply response may also lead for-profit suppliers to respond to the incentives for engaging in fraud.

This dilemma has characterized recent U.S. experience with nursing homes. The not-for-profit homes run by religious organizations appear to provide relatively high quality services in an honest manner. However, these homes typically have long waiting lists and have not responded to excess demand by expanding or creating new facilities. Nursing homes that are for-profit institutions have responded to increased demand by expanding rapidly. However, the experience, notably in New York State, has been one of considerable regulatory concern with fraud and low quality care.[17]

The question of whether a tuition tax credit plan would provide net tax refunds for low income families that choose private education for their children provides another example of conflict between objectives. Without refunds, one of the most attractive aspects of tuition tax credits--providing new educational opportunities for families currently sorted out of high quality schools--would not be realized. However, refunds might increase the incidence of fraud. The reason is not that low income families without positive tax liabilities are less honest than families with higher incomes; rather, refunds create incentives for unscrupulous parties to file refund claims on behalf of fictitious low income families. Recent U.S. experiences with food stamps indicate that such practices exist, and that the principal perpetrators are not the poor, but well organized, affluent criminals.

The analysis in Part IV emphasizes two points central to the policy debate over tuition tax credits. First, the impact of tuition tax credits on education in the United States would depend to a large extent on the details of the regulations defining the program. It is these regulations that would determine how tuition tax credits would influence access to private schools, the quality of education provided by private and public schools, the cost of education in these schools, and the incidence of fraud. Discussions of the consequences of tuition tax credits that do not pay attention to the details of the regulations defining the system are likely to mislead and confuse the policy debate. The second point is that analysis of alternative regulations should focus on their impact on the way sorting takes place in the private and public sectors, and on the incentives provided by alternative third party payment plans.

Footnotes

1. See Hanushek (1979) and Murnane (1981b) for reviews of the school effectiveness literature.
2. The evidence is actually a positive relationship between school district per pupil expenditures and prices of houses with comparable physical characteristics.
3. One recent study (Children's Defense Fund, 1974) reports that 8 percent of the students in U.S. public high schools were suspended at least once during the 1972-73 school year.
4. The prevalence of sorting in the public sector weakens the argument (Ackerman, 1980) that all students should be educated in public schools in order to assure that they are taught a common set of values. There is evidence (Silberman, 1970) that the style and substance of teaching in schools primarily serving middle class children is very different from that of schools serving children from poor families.
5. In calculating the percentage of the variation in student SES within each sector that consists of between school variation, the design weights were applied to correct for the oversampling of certain types of schools. A description of the methodology used to do the ANOVA calculations using weighted data is available from the author.
6. Many public school educators argue that most students dismissed from private schools end up in the public sector. If the HSB survey staff is successful in tracing students over time who do change schools, then the second wave of the longitudinal HSB survey should provide data to examine the extent and nature of student transfers.
7. The sensitivity of the results to methodological decisions is

discussed in the appendix.

8. To understand why a student who does not have attributes valued by other families would have limited educational options in a regime of competitive markets, think about the incentives for the parents of a student who does have valued attributes, and consequently is viewed as an asset to his or her classmates. This question can be framed in terms of how the parents could be compensated for the benefits such a child bestows on classmates. One way such compensation could take place is that the parents of the classmates might agree to subsidize the valued child's education by providing a scholarship. Some of this clearly takes place. Note, however, that a system of scholarships for students who are particularly valuable classmates requires much more complicated transactions than those associated with simple markets in which all consumers face the same prices. A second method of compensation is that the parents could enroll their students in a school that accepted only students who provided positive benefits to other students. Thus the family would be compensated by the positive benefits bestowed on their child by other students.

   Hamilton (1975) provides a formal presentation of equilibrium conditions for a problem very similar in structure to the sorting problem in education--namely, how people will group themselves into communities. He shows that communities will tend to specialize in serving families with particular attributes.

9. The percentage of lay teachers in Catholic elementary and secondary schools has risen from 28 percent in 1960 to 69 percent in 1979 (The Official Catholic Directory).

10. For Catholic schools operating at less than full capacity, the marginal cost of educating an additional student is probably significantly below the average cost, and possibly below existing tuition levels as well. For schools in this position, tuition tax credits may ease the financial constraints by increasing the number of students willing to pay the existing tuition. This may aid Catholic elementary schools, many of which have experienced declining enrollments over the last 20 years. (As indicated in Table 4 below, the average number of students per school in parish and diocesan elementary schools has fallen from 424 students in 1960 to 292 students in 1979.)

Catholic high schools (and elementary schools run by religious orders) have experienced a different enrollment pattern over the last 20 years. As shown in Table 4, although the number of Catholic high schools has declined over the past 20 years, the average enrollment per school has increased in both high schools run by parishes and dioceses, and in high schools run by religious orders. This increase is the result of the closing of a disproportionate number of small schools, and the movement to the larger Catholic schools of many students who would have attended the smaller schools. This consolidation policy saves resources, in part, by eliminating excess capacity and increasing utilization rates in the schools that do remain open. To the extent that these policies have resulted in the larger schools operating at near capacity, fiscal problems can be eased only by higher tuitions, not simply by admitting more students at existing tuitions.

11. These data come from the 1977 Census of Service Industries.

Table 4

Enrollment patterns in Catholic schools in the U.S., 1960 and 1979

	High schools run by parishes and dioceses		High schools run by religious orders		Elementary schools run by parishes and dioceses		Elementary schools ru by religious orders	
	Number of schools	Average enrollment per school	Number of schools	Average enrollment per school	Number of schools	Average enrollment per school	Number of schools	Average enrollment per school
1960	1567	332	866	374	9897	424	475	190
1979	905	566	637	536	7929	292	311	214

Source: The Official Catholic Directory, 1970 through 1979.

12. One way to make the sorted out students more attractive to profit making schools is to provide private schools with additional income. for educating such students. This has been suggested by Jencks and his colleagues in a volume entitled Education Vouchers (1970).

13. Research by Dennis Young (1980) suggests that there is considerable variation in the motivations of entrepreneurs in the not-for-profit sector. Consequently, one would expect considerable variation among responses to an increased demand from low income families. However, not enough is known yet about not-for-profit schools to predict whether the schools that would expand are schools that supply high quality services.

14. My analysis of the HSB data revealed that, after controlling for the background characteristics of individual students, the other private school sector included the schools in which students had the highest average achievement and schools in which the students had the lowest average achievement of all schools in the HSB sample, including public schools.

15. Designing regulations that uphold minimum quality standards without resulting in significant cost increases has been a problem in many human service industries--which typically are characterized by a lack of well defined input-output relationships. For example, see Young and Finch, 1977, pp. 221-238.

16. The incidence of fraud may be less under tuition tax credit plans than under voucher plans, in which payments are made directly from governments to suppliers of services.

17. See "Regulating Nursing Home Care: The Paper Tigers," Report of the New York State Moreland Commission, October, 1975.

APPENDIX

Notes on the methodology used to generate the predicted test scores and standard errors in Table 3.

There are a number of small differences between my methodology and CHK's.

1. CHK pooled the Catholic school and other private school subsamples in estimating equation (1). (Their equation included separate intercepts for the two subsamples.) The reason they did not pool the private school subsamples with the public school subsample and conduct a simple analysis of covariance was that the coefficients on the background variables differed for the public school and private school subsamples. My F test results indicate that the coefficients on the background variables for the Catholic school subsample are different from those for the other private school subsample. Consequently, extending CHK's logic, I estimated equation (1) separately for the three subsamples.

   It is important to keep in mind that the other private school subsample is extremely small (361 students in 21 schools). Moreover, the limited evidence available in the HSB data indicates that schools in this sector are extremely diverse—in tuitions, in academic programs, and in student achievement. Consequently very little can be learned from the HSB data about schools in this sector.

2. The HSB sample is a stratified sample that oversamples students in certain types of schools, including public and private schools attended by large numbers of minority students. In estimating

equation (1) CHK weighted their observations by the design weights. I used unweighted data on individual students in order to preserve the homoscedastic property of the error terms. However, CHK and I both used the design weights in calculating the average characteristics of high school seniors attending public schools in the U.S. These characteristics were then attributed to the hypothetical average public school senior whose performance was examined in the two experiments.

The results of the two experiments are somewhat sensitive to the choice of ols or weighted least squares in estimating equations (1) and (2). When the experiments are based on the weighted least squares estimates (CHK's procedure), 25 percent of the advantage of Catholic schools over public schools is estimated to be the result of student body composition effects (as opposed to 60 percent when the experiments are based on the ols estimates). The results of the public school-other private school comparison are not sensitive to the choice of estimation technique. Using either technique, the predicted test scores indicate that all of the advantage of other private schools over public schools is due to student body composition effects.

3. CHK used 17 background variables in estimating equation (1). I used only 10 in order to minimize the missing data problem that led CHK to employ the method of pairwise deletion of missing data—a method that many statisticians find troubling.

4. Unlike CHK, I calculated the standard errors appropriate for testing whether the achievement of the average public school student would be different if he or she attended a Catholic or other private school. A description of the method used to calculate these standard errors is

available from the author on request.

5. A final methodological issue concerns the specification of equation (2), which does not contain variables describing the quality of teachers and school programs--both determinants of school quality. If these variables are correlated with the average SES of the students in the school, their influence will be attributed to SES. It can be argued that this is not a specification error since a school's ability to attract high quality teachers and to implement effective homework and discipline practices depends on the composition of the student body. However, it would have been desirable to investigate whether the predictions for the second hypothetical experiment would be different if equation (2) included teacher and program characteristics. This was not done for two reasons. First, the HSB data set includes no information on the characteristics of teachers that have been found to be related to teaching performance--variables such as verbal ability and the quality of the teacher's undergraduate college. Second, the information in the HSB dataset on homework and discipline refer to the policy outcomes, not to the policies themselves. Since the outcomes (what the disciplinary environment is like and how much homework is completed) are endogenous, they do not belong on the right-hand side of equation (2) unless this equation is embedded in a larger system that includes equations explaining the determinants of these outcomes.

References

Ackerman, B. *Social Justice in the Liberal State*. New Haven: Yale U. Press, 1980.

Bryk, A.S. Disciplined inquiry or policy argument? *Harvard Educational Review 51* (November 1981), 497-509.

Children's Defense Fund. *Children Out of School in America*. Washington, D.C.: Children's Defense Fund. Washington Research Projects, 1974.

Coleman, J., Campbell, E.Q., Hobson, C.F. et al. *Equality of Educational Opportunity*. Washington, D.C.: Office of Education, National Center for Education Statistics, 1966.

Coleman, J., Hoffer, T., and Kilgore, S. *Public and Private Schools*. Chicago: National Opinion Research Center, 1981a.

Coleman, J., Hoffer, T., and Kilgore, S. Questions and answers: our our response. *Harvard Educational Review 51* (November 1981b), 526-546.

Edel, M., and Sklar, E. Taxes, spending and property values: Supply adjustment in a Tiebout-Oates model. *Journal of Political Economy* (September/October 1974).

Education Vouchers. A Report on Financing Elementary Education by Grants to Parents. Cambridge, Mass.: Center for the Study of Public Policy, 1970.

Fetters, W., Owings, F., Peng, S., and Takai, R. Review of NORC Report, Public and Private Schools, National Center for Education Statistics Memorandum, June 26, 1981.

Friedman, M. *Capitalism and Freedom*. Chicago: U. of Chicago Press, 1962.

Fuerst, J.S. Report card: Chicago's all-black schools. *Public Interest 64* (Summer 1981).

Goldberger, A.S., and Cain, G.G. The causal analysis of cognitive outcomes in the Coleman, Hoffer and Kilgore Report. Sociology of Education 55 (April/July 1982).

Grether, D., and Mieszkowski, P. Determinants of real estate values. Journal of Urban Economics (April 1974).

Hamilton, B. Zoning and property taxation in a system of local governments. Urban Studies 12 (1975).

Hanushek, E.A. Conceptual and empirical issues in the estimation of educational production functions. Journal of Human Resources 14 (Summer 1979).

Hanushek, E.A. Education and Race. Lexington, Mass.: D.C. Heath, 1972.

Hanushek, E.A. Throwing money at schools. Journal of Policy Analysis and Management 1 (Fall 1981), 19-41.

Henderson, V., Mieszkowski, P., and Sauvageau, Y. Peer group effects and educational production functions. Journal of Public Economics 10 (1978), 97-106.

James, E. Some reflections on cross-subsidization by NPO's. Mimeo, n.d.

King, A.T. Estimating property tax capitalization: A critical comment. Journal of Political Economy 85 (April 1977), 425-431.

Murnane, R.J. Evidence, analysis and unanswered questions. Harvard Educational Review 51 (November 1981a), 483-489.

Murnane, R.J. Interpreting the evidence on school effectiveness. Teachers College Record 83 (Fall 1981b), 19-35.

Nelson, R.R. Assessing private enterprise: an exegesis of tangled doctrine. Bell Journal of Economics 12 (Spring 1981), 93-111.

Noell, J. Public and Catholic schools: A reanalysis of "Public and Private Schools." Sociology of Education 55 (April/July 1982).

Oates, W. The effects of property taxes and local public spending on property values, a comment and yet further results. Journal of Political Economy (July/August 1973).

Oates, W. The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the Tiebout hypothesis. Journal of Political Economy (November/December 1969).

The Official Catholic Directory, 1970 through 1979. New York: P.J. Kennedy and Sons.

Page, E.B., and Keith, T.Z. Effects of U.S. private schools: A technical analysis of two recent claims. Educational Researcher 10 (August/September 1981), 7-17.

Pollakowski, H. The effects of property taxes and local public spending on property values: A comment and further results. Journal of Political Economy (July/August 1973).

Regulating Nursing Home Care: The Paper Tigers. Report of the New York State Moreland Commission, October, 1975.

Reinhard, R.M. Estimating property tax capitalization: A further comment. Journal of Political Economy 89 (December 1981), 1251-1260.

Rosenbaum, J.E. Making Inequality: The Hidden Curriculum of High School Tracking. New York: Wiley/Interscience, 1976.

Rosenbaum, J.E. Social implications of educational grouping. Chapter 8 in D.C. Berliner (ed.), Review of Research in Education 8. American Educational Research Association, 1980.

Silberman, C.E. Crisis in the Classroom. New York: Vintage Books, 1970.

Summers, A.A., and Wolfe, B.L. Do schools make a difference. American

Economic Review 67 (1977).

U.S. Department of Commerce, Bureau of the Census. 1977 Census of Service Industries, Other Service Industries, Parts 1-4. Geographic Area Studies (SC77-A-53).

West, E.G. Choice or monopoly in education. Policy Review 15 (Winter 1981), 103-117.

Willms, D. Achievement outcomes in public and private schools: A closer look at Coleman and NAEP data. Working Paper. Stanford: Institute for Research on Educational Finance and Governance, Stanford University (October 1981).

Winkler, D.R. Educational achievement and school peer group composition. Journal of Human Resources 10 (1975), 189-204.

Young, D. Entrepreneurship and the behavior of nonprofit organizations: Elements of a theory. Working Paper. New Haven: Program on Nonprofit Organizations, Institution for Social and Policy Studies, Yale University (July 1980).

Young, D., and Finch, S. Foster Care and Nonprofit Agencies. Lexington, Mass.: D.C. Heath, 1977.

Comparing Public and Private Schools:
The Puzzling Role of Selectivity Bias

by

Richard J. Murnane
Stuart Newstead
Randall J. Olsen

Institution for Social and Policy Studies
Yale University

September, 1982

The research for this paper was supported by Grant NIE-G-79-0084 from the National Institution of Education, and by grants from the Institution for Social and Policy Studies and the Program on Nonprofit Organizations, Yale University.

INTRODUCTION

Research on the controversial question of whether private schools in the United States are more effective than public schools in enhancing student achievement has been hindered by a variety of conceptual problems and data limitations. At the center of the research difficulties is the problem of distinguishing student achievement differences due to the effectiveness of school programs from those due to student abilities. This problem is particularly difficult because the school choices made by American families, who are faced with varied schooling alternatives and differing financial constraints, result in significant selection of students with particular backgrounds and abilities to particular schools. Unless the influences on student achievement of student attributes are controlled, the estimates of school program effects will be contaminated by what is known in the econometric literature as selectivity bias. It is now well known that the conventional method for controlling the effects on student achievement of the attributes of students attending different schools--the inclusion of variables describing student backgrounds in a single equation, multiple regression framework--does not, in general, eliminate selectivity bias, and consequently does not produce reliable comparisons of the relative quality of public and private schools (Barnow et al., 1980).

In recent years new techniques have been developed to deal with selectivity bias that in principle could be helpful in developing relaible estimates of the relative effectiveness of public and private schools. Among the contributors to this new methodology are Goldberger (1972, 1980), Gronau (1973, 1974), Maddala and Lee (1976), Olsen (1980, 1982), and, most importantly, Heckman (1974, 1976, 1978, 1979). These techniques have quickly come into widespread use in evaluating education and manpower training programs (Farkas et al., 1980; Mallar et al., 1980), and in estimating demand equations (McGuire, 1980; Willis and Rosen, 1979) and production functions (Orazem, forthcoming). For 1981 alone, the Social Science Citation Index lists 79 references to the Heckman article.

Until 1981, a lack of data prevented the application of the new techniques for controlling selectivity bias to the question of the relative effectiveness of public and private schools. However, in that year a large data set became available that provides information on the backgrounds and skill levels of large numbers of students attending public and private (predominantly Catholic) high schools. To this date, two sets of papers have applied the new techniques to the new data. The results have not clarified the relative quality issue, however. In fact, the studies report conflicting estimates of the relative effectiveness of public and Catholic high schools. Since both sets of papers were based on the same data and both used variants of the new techniques for controlling selectivity bias, the conflict between the results poses a significant puzzle.

This research was undertaken to solve the puzzle of the conflicting results. As the research progressed, a second theme developed--namely, that the results of applying the new techniques for controlling selectivity bias can be extremely sensitive to a number of assumptions, and consequently it is important to adopt an analysis strategy that permits investigation of these assumptions.

To the reader interested only in the substantive puzzle or only in selectivity bias methodology, the organization of this paper may be initially frustrating in that the two themes are interwoven. However, we believe that this is necessary for two reasons. First, understanding the solution to the substantive puzzle requires a thorough understanding of the new methodology for controlling selectivity bias and the different ways in which this methodology can be applied. Second, the many assumptions involved in applying the selectivity bias methodology and the methods that can be used to investigate the validity of the assumptions can best be explained in the context of a substantive problem.

II. THE PUZZLE

In April 1981, Coleman, Hoffer, and Kilgore (henceforth CHK) completed a highly publicized study of the relative effectiveness of public and private high schools in the U.S. Their report, entitled Public and Private Schools, used the baseline data from High School and Beyond (HSB), a federally funded longitudinal study of students who were in their sophomore or senior year in a U.S. high school in 1980. Eighty-seven percent of the students in the sample attended public schools, nine percent attended Catholic schools, and three percent attended other non-Catholic private schools. Since the non-Catholic private schools represented in the sample formed a very small yet exceedingly diverse group, attention has focused on differences between public schools and Catholic schools, and this paper will address only the reported public-Catholic school comparisons.

The most controversial aspect of CHK's report was the conclusion that Catholic schools are more effective than public schools in enhancing the cognitive skills of students (as measured by scores on tests of reading and mathematics). Critics attacked many aspects of the report, but perhaps the most common criticism concerned the methodology used to generate the public school-Catholic school achievement comparisons. CHK attempted to control for differences between the attributes of public and Catholic school students by including 17 background variables in equations predicting student achievement. These equations were estimated using ordinary least squares. (See Goldberger and Cain, 1982, for a detailed description of CHK's original methodology.)

The basic criticism of this methodology was that even the inclusion of a large number of family background variables in an equation predicting student achievement does not necessarily eliminate selectivity bias (Barnow et al., 1980). This criticism raised the issue of whether alternative techniques, such as those developed by Heckman, were appropriate for examining differences in the effectiveness of public and Catholic schools and whether the use of such techniques would produce different results.

Later in 1981 and again in 1982, papers by Noell based on the HSB data reported that the results were different when public-Catholic school differences were estimated in a framework that explicitly modelled the selection process. Noell's results, based on estimation of a Heckman-type model, indicated that, contrary to CHK's original results, there were almost no statistically significant differences between the effectiveness of public and Catholic schools in producing cognitive skills in their students.

CHK responded to the criticisms of their ordinary least squares methodology by also bringing the Heckman technique to bear on the HSB data. However, they reported that this produced a larger estimated advantage of Catholic schools over public schools than ordinary least squares (CHK, 1981b: 529-30).

By themselves, neither the CHK nor the Noell results are illogical, since the direction of bias in the estimated program effect produced by ordinary least squares is not known a priori (Barnow et al., 1980). However, the conflict between the results of studies that apply the same general estimation strategy to the same database is puzzling.

III. SUMMARY OF THE TWO STEP TECHNIQUES FOR CONTROLLING SELECTIVITY BIAS

A. General Framework

Section IIIA, which is based on Barnow, Cain and Goldberger (1980), provides a brief formal description of the selectivity bias problem in the context of public school-Catholic school comparisons. Sections III B and III C describe and compare the variants of the new methodology for controlling selectivity bias adopted by Noell and CHK.

For the ith child, $(i=1,\ldots n)$, let

y_i = test score

z_i = school type (1=Catholic, 0=public)

X_{1i} = exogenous vector of k background variables, including 1

X_{2i} = exogenous vector of m background variables, where X_{1i} is a subset of X_{2i}

t_i = unobserved continuous variable determining school type

The model which underlies Heckman's method for controlling for selectivity bias contains an outcome equation (3.1) and a selection equation (3.2):

$$y_i = X_{1i}{}'\beta_1 + \alpha z_i + u_{1i} \tag{3.1}$$

$$t_i = X_{2i}{}'\beta_2 + u_{2i} \tag{3.2}$$

$$z_i = \begin{cases} 1, & \text{if } t_i \geq 0 \\ 0, & \text{if } t_i < 0 \end{cases}$$

u_{2i} is normally distributed and $E(u_{1i}/u_{2i})$ is a linear function of u_{2i} such that:

$$E(u_{1i}) = E(u_{2i}) = 0$$

$$\text{var}(u_{1i}) = \sigma_1^2\,; \quad \text{var}(u_{2i}) = 1$$

$$\text{cov}(u_{1i},u_{2i}) = \rho\sigma_1\,;$$

$$\text{cov}(u_{1i},u_{1j}) = \text{cov}(u_{2i},u_{2j}) = \text{cov}(u_{1i},u_{2j}) = 0 \text{ if } i \neq j.$$

The standardization of u_{2i} to have unit variance entails no loss of generality.

Now define $\theta_i = -X_{2i}'\beta_2$. Barnow et al. show that

$$E(u_{2i}/\theta_i,z_i) = z_i\, f(\theta_i)/(1-F(\theta_i)) - (1-z_i)f(\theta_i)/F(\theta_i) \tag{3.3}$$

$$= h_i(\theta_i,z_i) \text{ , say.}$$

$f(\cdot)$ and $F(\cdot)$ represent the standard normal density and distribution functions. For notational convenience, let $h_i = h_i(\theta_i,z_i)$.

It follows that

$$E(u_{1i}/\theta_i,z_i) = \rho\sigma_1 h_i \tag{3.4}$$

Equation (3.4) shows that ordinary least squares applied to equation (3.1) will lead to biased estimates of the parameters α and β_1 unless $\rho=0$. h_i is positive if $z_i=1$ and negative if $z_i=0$, so the Catholic school advantage, α , will be overestimated if ρ is positive, and underestimated if ρ is negative. In other words, the Catholic school advantage will be overestimated (underestimated), if more able (less able) students of a given family background have a tendency to choose Catholic schools over public schools.

B. Estimation Under the Assumption of One Student Population

If we assume that the student bodies of public and Catholic schools are drawn from a single population of students, all of whom attend either public or Catholic high schools, and for whom the values of β_1 are independent of school choice, then consistent estimates of the parameters can be derived by the following two step method:

First, use maximum likelihood probit analysis to estimate β_2 from the model

$$P[z_i=1] = F(X_{2i}'\beta_2) \tag{3.5}$$

Call this estimator $\hat{\beta}_2$ and employ either one of the following second steps:

(a) replace z_i in equation (3.1) by $\hat{z}_i = F(X_{2i}{}' \hat{\beta}_2)$, and estimate equation (3.6) by least squares,

$$y_i = X_{1i}{}' \beta_1 + \alpha \hat{z}_i + \varepsilon_i \tag{3.6}$$

or (b) calculate $\hat{\theta}_i = -X_{2i}{}' \hat{\beta}_2$,

and $\hat{h}_i = z_i f(\hat{\theta}_i)/(1-F(\hat{\theta}_i)) - (1-z_i)f(\hat{\theta}_i)/F(\hat{\theta}_i)$,

add the auxiliary regressor $\hat{h}_i$ to equation (3.1), let $c = \rho\sigma_1$, and estimate equation (3.7) by least squares,

$$y_i = X_{1i}{}' \beta_1 + \alpha z_i + c\hat{h}_i + \eta_i \tag{3.7}$$

Noell chose the first of the two estimation strategies.

Although the parameter estimates in equations (3.6) and (3.7) are consistent, the standard errors yielded by ordinary least squares are not, in general, correct because the errors ε_i and η_i are heteroscedastic unless $\rho=0$ (Heckman, 1976). Appendix A2(a) describes a simple method for obtaining correct standard errors on the estimates of α , β_1 , and c in equation (3.7).

C. Estimation Under the Assumption of Two Student Populations

The estimation strategy used by CHK is based on the premise that the structure of equation (3.7), including the values of β_1 and c , is different for the public and Catholic school student populations. Consistent estimates of the parameters, β_1 and c , can be derived for each of the two populations by extending the methodology of the one population model as follows:

First, as in the case of the one population model, run maximum likelihood probit on all observations and construct $\hat{h}_i$ for each student. Thus,

$\hat{h}_i = f(\hat{\theta}_i)/(1-F(\hat{\theta}_i))$ for the ith Catholic school student

$\hat{h}_j = -f(\hat{\theta}_j)/F(\hat{\theta}_j)$ for the jth public school student.

Second, use least squares to estimate

$y_i = X_{1i}'\beta_1^P + c^P \hat{h}_i + \eta_i^P$ (3.8) for the public school subsample

$y_j = X_{1j}'\beta_1^C + c^C \hat{h}_j + \eta_j^C$ (3.9) for the Catholic school subsample.

Correct standard errors can be calculated by a method very similar to that used in the one population case.

From the estimates of equations (3.8) and (3.9), an estimate of the Catholic school advantage for a student with a particular k-vector of characteristics, x^*, can be calculated as

$\hat{y}^C - \hat{y}^P = x^{*\prime}\beta_1^C - x^{*\prime}\beta_1^P$, with standard error, $[x^{*\prime}(V^P+V^C)x^*]^{\frac{1}{2}}$

where

V^P = the variance-covariance matrix of the estimated β^P coefficients

V^C = the variance-covariance matrix of the estimated β^C coefficients.

In estimating the Catholic school advantage, CHK defined x^* to be the average characteristics of students attending public high schools.[1]

There are two differences between the one and two population models that are important to recognize. First, since β^C and β^P are not constrained to be equal, the estimate of the Catholic school advantage may depend critically on x^*; second, since c^P and c^C are not constrained to be equal, a finding of "cream skimming" in the Catholic school sample ($c^c > 0$) need not imply "bottom scraping" ($c_p > 0$) in the public school sample. This would be implausible if all students of high school age attended either public or Catholic schools, but some teenagers choose non-Catholic private schools and others choose not to attend school at all. As a result, the nature and extent of student selection in the public and Catholic schools could be different.

As we show below, the choice of a one population or two population model does play a role in explaining the difference between Noell's and CHK's results. The differences between the one and two population models are emphasized here because many articles in the evaluation literature that discuss the application of the new techniques for controlling selectivity bias do not clarify the implicit assumptions involved in the choice of the one or two population model (e.g., Barnow et al., 1980).

IV. OUR RESEARCH STRATEGY

A. General Framework

One problem with the strategies used by CHK and Noell is that estimation of the selection equation (3.5) by probit analysis is computationally expensive. Since our basic strategy for unravelling the puzzle of the conflicting results was to examine the sensitivity of the results to the many small differences distinguishing the two methodologies, it was important to adopt a low cost estimation strategy. It is possible to develop computationally inexpensive techniques by assuming that instead of being normally distributed, u_{2i} is distributed uniformly over the interval $[0,1]$ for each $i=1,\ldots,n$.

B. Estimation Under the Assumption of One Student Population

As shown in Appendix A2(b), consistent estimates of the parameters of equation (3.1) for the one population model can be derived by the following two step method.

First, estimate $\hat{z}_i$, the probability that z_i is 1, for each observation, using the linear probability model:

$$\hat{z}_i = P[z_i=1] = X_{2i}'\tilde{\beta}_2 . \tag{4.1}$$

Second, add the auxiliary regressor $\hat{s}_i = z_i - \hat{z}_i$ to equation (3.1) and estimate equation (4.2) by least squares,

$$y_i = X_{1i}'\beta_1 + \alpha z + \delta\hat{s}_i + \nu_i \tag{4.2}$$

As with equations (3.6) and (3.7), standard errors on the estimates of β_1, α, and δ will, in general, be incorrect if ordinary least squares is applied to equation (4.2). Appendix A2(b) also shows how correct standard errors may be derived.

This technique, which we have called the "s method," yields consistent estimates of the coefficients. Appendix A3 shows that the estimated coefficients produced by this method are identical to the estimates provided by two stage least squares, in which the first stage consists of estimation of the linear probability model, equation (4.1). The s method has the advantage, however, of providing a direct test of the null hypothesis of no selectivity bias $(\mathrm{cov}(u_{1i},u_{2i})=0)$. The null hypothesis will be rejected if the estimate of δ is significantly different from zero, when compared to its standard error.

C. Estimation Under the Assumption of Two Student Populations

The only difference between our method for estimating the Catholic school advantage and the method developed by Heckman and used by CHK is that we estimate the selection equation with ordinary least squares instead of probit analysis. Thus, we replace $\hat{h}_i$ in equations (3.8) and (3.9) with $\hat{s}_i$ and estimate equation (4.3) for the public school subsample and equation (4.4) for the Catholic school subsample:

$$y_i = X_{1i}'\beta_1^p + \delta^p \hat{s}_i + \nu_i^p \tag{4.3}$$

$$y_j = X_{1j}'\beta_1^c + \delta^c \hat{s}_j + \nu_j^c \tag{4.4}$$

D. Comparison of Least Squares and Probit Methods

We found that the predicted probabilities of Catholic school attendance generated by least squares and probit methods were very close—the correlation coefficient exceeded .99. As would be expected under these circumstances, the estimates of the Catholic school advantage generated by these alternative methods were very similar. Thus, our methods provide a low cost strategy for examining the sources of the puzzle posed by the conflict between CHK's and Noell's results.

The s method requires an exclusion restriction to identify the achievement equations—(4.2) in the one population model and (4.3) and (4.4) in the two population model. However, this is not a serious disadvantage of the s method relative to the probit based methods, since application of the latter often produces unstable results when identification is made solely through functional form (Olsen, 1980). CHK reported this instability when they applied the Heckman methodology to the HSB data with no exclusion restriction (CHK, 1981b: 529-30). In further probit based analyses, both

CHK and Noell adopted the identifying restriction that a student's religious status (1=Catholic, 0=other) influences choice of school type, so is included in the X_2 vector, but does not influence achievement, so is not included in the X_1 vector. We chose the same identifying restriction.

For reasons of economy, we conducted our work with a subsample of 5500 observations from the HSB sample: all sophomores in Catholic schools for whom complete data were available plus all sophomores with complete data in a random sample of 125 of the 988 public schools in the survey.

V. WHY CHK'S AND NOELL'S RESULTS DIFFER

A. Two Reasons

There are many differences between the specifications of CHK's and Noell's models, including the choice of background variables in the achievement and selection equations (X_{1i} and X_{2i}), and the choice of scale used to measure the dependent variable. These differences influence the results to some extent. However, our sensitivity analysis indicates that none of these differences accounts for the primary conflict. Instead, the difference stems from the following two factors, listed in decreasing order of importance:

1. different choices about whether to weight each student in the sample equally in estimating the model, or whether to use the design weights to give some students greater weight than others,
2. different choices about the use of a one population model or a two population model.

B. Results Using a One Population Model with Different Weighting Options

The HSB sample is a stratified random sample, with an oversampling of students in certain types of schools, including public and Catholic schools in which at least 30 percent of the students are minority group members. The database includes design weights that in principle permit the creation of a weighted sample that reflects the U.S. high school student population in 1980. One must therefore decide whether to give each student in the sample equal weight in estimating the model of selection and achievement, or whether to use the design weights to give some students greater weight than others. CHK chose the first option and did not apply the design weights;[2] Noell chose the second option.

The importance of the weighting decision can be seen by examining the first row of Table 1. This row provides estimates of the Catholic school advantage using both weighted and unweighted data.[3] These estimates were made using the s method and a one population model. In these estimations, all of the many small differences between Noell's and CHK's specifications of X_1 and X_2 were eliminated. A set of eight background variables, plus a constant, constituted X_1 in all of the estimations reported in Table 1. These eight variables plus a constant and the variable indicating Catholic religious status constituted X_2. Summary statistics describing the distributions of all variables are provided in appendix A1.

Table 1 Here

When the model is estimated using the unweighted data, the estimated Catholic school advantage in imparting reading skills is more than twice the estimated advantage when the model is estimated using weighted data. Moreover, with unweighted data, the coefficient is significantly different from zero; with weighted data, it is not. Weighting also affects the estimates of the Catholic school advantage in imparting mathematics skills.

Table 1

Estimates of the Differences in the Reading and Math Skills of Students Attending Catholic and Public Schools, Using a One Population Model With Observations Pooled Across Ethnic Groups[a]

	Reading		Math	
	Unwgted Data	Wgted Data	Unwgted Data	Wgted Data
1. The Catholic school advantage (α) from estimating equation (4.2) with the s method	0.65* (0.21)	0.25 (0.67)	1.53* (0.38)	1.59 (1.22)
2. The extent of selectivity bias (δ)	0.07 (0.24)	0.43 (0.70)	-0.40 (0.44)	-0.61 (1.28)
3. R^2 from estimating equation (4.2) with the s method	.14	†	.19	†
4. The Catholic school advantage (α) obtained by estimating equation (3.1) by ordinary least squares	0.70* (0.10)	0.65* (0.20)	1.23* (0.19)	1.04* (0.36)
5. R^2 from estimating equation (3.1) by ordinary least squares	.14	†	.19	†

Standard error in parentheses

[a] The complete regression results are available upon request from the first author.

* statistically significant on a 2-tailed 5% t test.

† Since there is no intercept when the achievement equations are estimated with weighted data, the R^2 is not a meaningful statistic.

In this case, the estimated coefficients are close in size, but only the coefficient estimated with the unweighted data is significantly different from zero. Thus, when the model is estimated with weighted data, the conclusion is that there are no significant differences in the effectiveness of Catholic and public schools. When the model is estimated with unweighted data, the conclusion is that Catholic schools are more effective than public schools in teaching reading and mathematics.

It is important to emphasize that the differences in results stem not simply from the weighting choices independent of model specification. The fourth row of Table 1 shows the Catholic school advantages from estimating equation (3.1) by ordinary least squares, with weighted and unweighted data. The results based on weighted and unweighted data suggest the same qualitative conclusions--namely, statistically significant Catholic school advantages in teaching reading and math. Thus, the story is not simply the importance of the weighting decision, independent of model specification; rather, the importance of weighting depends on the specification of the particular model.

Why does the weighting decision affect the Catholic school advantage estimated in a model that explicitly deals with selectivity bias? We know that if the model were specified correctly, both weighted and unweighted data would provide unbiased estimates of the parameters; weighting would only influence the efficiency of the estimates. The sensitivity of the results to the weighting decision suggested that the model was misspecified.

One likely candidate for the source of misspecification was a difference in the structure of the model for different ethnic groups. If the structure of the model did differ across ethnic groups, then the results using a sample pooled across ethnic groups would be sensitive to the representation given to each group in the total sample. In the unweighted HSB sample, 12 percent of the students were classified black and 16 percent hispanic. In the weighted sample, black students constituted 10 percent of the sample and hispanic students eight percent. Hispanic and, to a lesser extent, black students, thus have greater influence when the model is estimated with unweighted data pooled across ethnic groups than when the model is estimated with weighted data.[4/]

We used the s method separately for white, black and hispanic students, employing both reading scores and mathematics scores as dependent variables in the achievement equation (4.2). The F test results, which are reported in the last row of Table 2, indicate that there are significant differences across ethnic groups. Most important, the estimates of the Catholic school advantage, also presented in Table 2, differ across ethnic groups. For white students there are no significant differences between Catholic and public schools in the reading and math scores. However, for black and hispanic students, the results indicate significant Catholic school advantages.

Table 2 Here

The variation across ethnic groups in the estimated Catholic school advantage provides a basis for explaining the conflict between CHK's and Noell's results. CHK did not use the sample weights in estimating the size of the Catholic school advantage, while Noell did. As a result, CHK's method gave greater weight to the statistically significant Catholic school advantage for black and hispanic students, which led them to draw a different conclusion from Noell about the relative efficacy of public and Catholic schools.

Table 2

Estimates of the Differences in the Reading and Math Skills of Students Attending Catholic and Public Schools, Using a One Population Model With Unweighted Data Stratified by Ethnic Groups[a]

	Reading			Math		
	White	Black	Hispanic	White	Black	Hispanic
1. The Catholic school advantage (α) derived from estimating equation (4.2) with the s method	0.20 (0.22)	2.38* (0.75)	3.04* (0.96)	0.42 (0.40)	4.99* (1.28)	8.01* (1.81)
2. The extent of selectivity bias (δ)	0.42 (0.26)	-0.99 (0.82)	-1.75 (1.00)	0.75 (0.48)	-3.04* (1.39)	-5.65* (1.86)
3. R^2 from estimating equation (4.1) with the s method	.10	.10	.13	.13	.08	.15
The Catholic school advantage (α) obtained by estimating equation (3.1) with ordinary least squares	0.49* (0.12)	1.55* (0.31)	1.42* (0.27)	0.93* (0.23)	2.47* (0.52)	2.81* (0.47)
5. R^2 from estimating equation (3.1) by ordinary least squares	.09	.10	.12	.13	.07	.14
6. F (16,5473) statistic from testing null hypothesis that the structure of equation (4.2) is the same for all ethnic groups		3.07*			5.81*	

Standard errors in parentheses

[a] The complete regression results are available upon request from the first author.

* statistically significant at 5 percent level.

Given the conflict between CHK's and Noell's results, the reader may ask: Which method is correct? The answer is that neither is correct. Both sets of results are flawed because pooling observations across ethnic groups conceals important differences among ethnic groups in the model's structure. The correct strategy is to estimate the model separately for each ethnic group, and to use unweighted data to preserve the homoscedastic properties of the error terms. (If the error terms for the unweighted sample are homoscedastic, then the application of the design weights will introduce heteroscedasticity and lead to incorrect estimates of the standard errors of the parameter estimates.)

C. Results for Different Ethnic Groups Using a Two Population Model

A second source of explanation for the conflict between CHK's and Noell's results is the choice of a one or two population model. Table 3 presents estimates of the Catholic school advantage for each ethnic group based on the two population model. The estimates were calculated for two sets of values of x^*: the average characteristics of students attending public schools and the average characteristics of students attending Catholic schools. (The design weights were used in calculating the appropriate means, but not in estimating the equation.) The estimates of the Catholic school advantage for black students are included for completeness, but are extremely unstable due to the low explanatory power of the predicting equations, particularly the public school equations.

Table 3 Here

The results in Table 3 illustrate two points. First, the estimates of the Catholic school advantage obtained with a two population model are quite

Ta[illegible] 3

Estimates of the Differences in the Reading and Math Scores of Students in Catholic and Public Schools, Using a Two Population Model with Unweighted Data Stratified by Ethnic Group

	Reading			Math		
	White	Black	Hispanic	White	Black	Hispanic
1. The Catholic school advantage, based on predicting the achievement of the average public school student in public schools (equation 4.3) and in Catholic schools (equation 4.4) (s.e. in parentheses)	0.28 (0.29)	2.40*† (0.79)	3.54* (1.33)	1.27* (0.52)	5.52*† (1.35)	9.46* (2.28)
2. The Catholic school advantage, based on predicting the achievement of the average Catholic school student in public schools (equation 4.3) and in Catholic schools (equation 4.4) (s.e. in parentheses)	0.07 (0.26)	2.53*† (0.93)	2.77* (1.01)	0.16 (0.47)	5.71*† (1.46)	7.44* (1.84)
3. Extent of selectivity bias in public school sample (δ^p) (s.e. in parentheses)	0.39 (0.30)	-1.20 (1.48)	-1.86 (1.04)	0.98 (0.55)	-3.85 (2.26)	-6.24* (2.04)
4. Extent of selectivity bias in Catholic school sample (δ^c) (s.e. in parentheses)	0.57 (0.57)	-0.84 (1.00)	-2.51 (2.32)	0.13 (1.04)	-2.82 (1.76)	-7.58* (3.84)
5. R^2 from predicting the achievement of the public school sample (equation 4.3)	.10	.02	.12	.15	.03	.14
6. R^2 from predicting the achievement of the Catholic school sample (equation 4.4)	.06	.04	.07	.07	.06	.09

* significant on a 2-tailed 5% t test.

† The estimates of the Catholic school advantage for black students are extremely unstable due to the low

sensitive to the choice of x^*. Second, for every ethnic group, the Catholic school advantage estimated from a two population model with x^* assuming the values of the characteristics of the average public school student is larger than the advantage estimated in a one population model (see Table 2). Thus, CHK's choice of the two population model and Noell's choice of the one population model contributed to the difference in their results.

VI. A NEW PUZZLE AND A PROPOSED SOLUTION

A. The New Puzzle

The results of estimating both the one and two population models for the separate ethnic groups indicate that the direction of selectivity bias is different for white students than for minority students. In both models, the estimated value of $\mathrm{cov}(u_1,u_2)$ is greater than zero for white students, although the coefficient is never large enough relative to its standard error to reject the null hypothesis of no selectivity bias. However, the estimates of $\mathrm{cov}(u_1,u_2)$ are negative for black and hispanic students. This implies that, among black and hispanic sophomores with the same background characteristics, students whose ability is lower are more likely to attend Catholic schools than public schools. The negative covariances, which are statistically significant for both black and hispanic students in the one population model and for hispanic students in the two population model estimated with math scores as the dependent variable, are somewhat counterintuitive. Moreover, it is not apparent why the direction of selectivity bias should be different for minority group students than for white students.

B. An Alternative Test of Selectivity Bias

In an attempt to solve this new puzzle, we applied an alternative technique for investigating selectivity bias (Olsen, 1982). Imagine that one knew what the population distribution of residuals from a regression of white students' test scores on their background characteristics would be if all white students attended Catholic schools. Compare this distribution with the distribution of residuals of white students who do attend Catholic schools. If there is a correlation between ability and choice of school type, then the Catholic school residuals will not be a random sample from the underlying population, and these two distributions will therefore have different shapes. In particular, if $\mathrm{cov}(u_1,u_2) > 0$, the observed Catholic school residuals will underrepresent the left tail of the population distribution.

Unfortunately, this strategy cannot usually be implemented directly because the distribution of the underlying population is rarely known; however, an approximation to the strategy is available. The achievement residuals of white Catholic school students who have a high estimated probability of attending a Catholic school, based on their background characteristics, should roughly represent the population distribution of whites' residuals. On the other hand, the achievement residuals of white Catholic school students who have a low estimated probability of attending a Catholic school, based on their background characteristics, should be affected by any selectivity bias that is present. If these two distributions are sufficiently different, this will be evidence of selectivity bias. For a more detailed exposition, see Olsen (1982).

We applied this technique to the samples of white, black, and hispanic students in public and Catholic schools, producing a total of six tests of selectivity bias. Math scores were used as the dependent variable in generating the distributions of residuals, since the differences across ethnic groups in the estimates of selectivity bias using the s method were more pronounced when math scores were used as the dependent variable than when reading scores were used. The results of the likelihood ratio tests are reported in Table 4.

Table 4 Here

One striking aspect of these results is that no selectivity bias was found among black or hispanic student samples in either public or Catholic schools. This is in contrast to the results obtained from the s method and reveals that the conclusion of significant selection of the less able minority group students to Catholic schools is an artifact of a specification error--namely, the assumption that Catholic religious affiliation does not influence student achievement. In fact, the significant coefficient, δ , on the auxiliary regressor, $\hat{s}$, actually reflects the influence of Catholic religious status on the achievement of minority group students.

To see this, compare equation (6.1) below to equations (4.1) and (4.2) that were estimated with the s method.

$$y_i = X_{1i}'\beta_1 + \alpha z_i + rR_i + e_i \tag{6.1}$$

Let R denote Catholic religious status and ξ be the coefficient on R in the selection equation (4.1). It is easily shown that

$$\delta = -r/\xi .$$

Table 4

$\chi^2(2)$ Statistics for Testing Null Hypotheses of No Selectivity Bias

* indicates significance at .05 level

	White	Black	Hispanic
Public	34.08*	0.12	0.18
Catholic	0.00	1.10	0.30
Numbers of observations in the low and high probability subsamples of students in:			
Public schools	2321†	322	423
Catholic schools	1671†	321	442

---

† To increase the power of the χ^2 statistic for the white students in each school sector, 1288 white students in public schools and 627 white students in Catholic schools who had a medium probability of being in a Catholic school were not included in any of the subsamples. The small sizes of the minority group samples of public and Catholic students made it necessary to include every observation in a subsample.

Since ξ is positive for all ethnic groups, δ assumes the opposite sign from r. As shown in Table 5, the estimates of r are negative for whites and positive for blacks and hispanics. This solves the puzzle posed in Section VI(A).

Table 5 Here

Thus, one lesson from the alternative test of selectivity bias is that the two step methods developed by Heckman and others, which examine whether the mean of the least squares residuals shifts with the probability of being in a particular sector, are sensitive to specification error. In the case of minority group students, these methods led to the inference that low ability students were selected into Catholic schools. The distributions of residuals for minority group students indicate that this inference is incorrect and that the negative values of δ stemmed from the improper exclusion restriction.

At face value, the exclusion restriction chosen by Noell and CHK appears reasonable. This only points out how difficult it is to properly specify models of human behavior and the importance of finding ways to test the validity of exclusion restrictions.

A second striking finding from the alternative test is that the results indicate significant selectivity bias among white students in public schools. The nature of this bias is illustrated by Figure 1, which shows the theoretical densities of the residuals for white students who have a high or low probability of being in a public school. The distribution for high

Figure 1 Here

Table 5

The Relationship Between Selectivity Bias and Catholic Religious Status[a]

	Pooled Data (n=5500)	Whites (3992)	Blacks (643)	Hispanics (865)
1. Coefficients on Catholic religious status (r) in equation (6.1) estimated by ordinary least squares	.20 (.22)	-.41 (.27)	1.15* (.52)	2.00* (.61)
2. Coefficient on Catholic religious status (ξ) in equation (4.1) estimated by the linear probability model	.50* (.01)	.55* (.01)	.38* (.03)	.35* (.04)
3. Extent of selectivity bias (δ) in equation (4.2) estimated by the s method	-.40 (.44)	.75 (.48)	-3.04* (1.39)	-5.65* (1.86)

Standard errors in parentheses

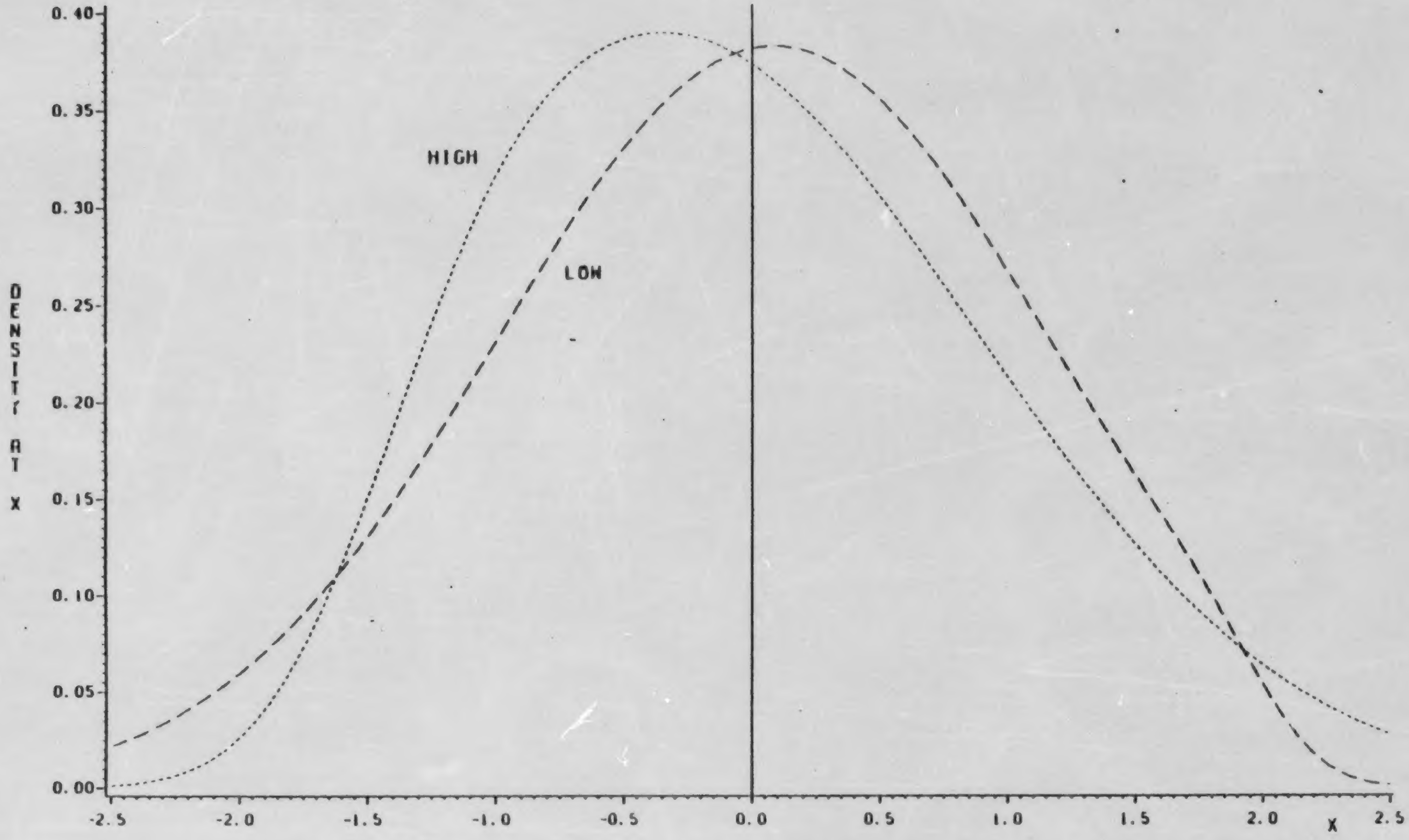
[a] The complete regression results are available upon request from the first author.

* statistically significant on a 2-tailed 5% t test.

The results reported here use unweighted data and math score as the dependent variable in rows 1 and 3.

Figure 1

ESTIMATED DISTRIBUTIONS OF RESIDUALS FROM LINEAR REGRESSIONS OF MATH SCORE ON ALL BACKGROUND VARIABLES (X_2)



WHITE STUDENTS IN PUBLIC SCHOOLS WITH A HIGH OR LOW ESTIMATED PROBABILITY OF ATTENDING A PUBLIC SCHOOL probability students is skewed to the left, implying either that the distribution of residuals for the underlying population of all white students is nonnormal, or that the sample distribution underrepresents low achieving students who are absent from school on test days. The estimated density of residuals for low probability public school students is skewed to the right relative to the high probability distribution, indicating underrepresentation of high achieving students. The chi-square test reveals that the distributions of the two sets of residuals are significantly different, indicating selectivity bias among white students in public schools.

The reason that the significant selectivity bias among white students in public schools did not show up when the s method was used is that, as was the case with minority group students, the achievement equation was unidentified and the effect of selectivity bias was confounded with the impact of religious status on student achievement. Thus, we see that an improper exclusion restriction can lead either to the conclusion of selectivity bias when there is in fact none, or to the conclusion of no selection when in fact selection is present.

It is interesting to note that the result of the chi-square test reveals no selectivity bias among white students in Catholic schools. This suggests that the bias among white students in public schools reflects the choices of high achieving students to attend non-Catholic private schools. It also emphasizes that the assumption implicit in the use of the one population model to compare the effectiveness of two programs, that "bottom scraping" in one program implies "cream skimming" in the other, should be examined carefully.

The presence of significant selectivity bias among white students in public schools means that the use of ordinary least squares will not produce consistent estimates of the relative efficacy of public and Catholic schools in educating white students. Is there an alternative methodology that will produce consistent estimates? The two step methods based on Heckman's and Olsen's work cannot be used unless an alternative exclusion restriction can be justified—and this is doubtful. If the density distribution of the residuals for white students in public schools (and Catholic schools) were known, then maximum likelihood methods could be used. However, the results presented above suggest that even the underlying distribution from which the sample is drawn may be nonnormal. Thus we can have little confidence in the maximum likelihood results. We do know, however, given the nature of the selectivity bias among white students in public schools, that the ordinary least squares estimates of the Catholic school advantage in equations that include family religious status provide an upper bound on the true advantage.

VII. LESSONS

Among the lessons to be derived from this paper that are relevant to the increasingly large number of users of the two step methods for dealing with selectivity bias are the following:

1. Selection can work differently for different subgroups in the population (for example, different ethnic groups). When this is the case, results can be extremely sensitive to the weights given to different groups in the sample used in empirical work.
2. In comparing the effectiveness of two alternative programs (for example, public and Catholic education) using a one population model, one makes the implicit assumption that nonrandom selection of participants to one program (e.g., cream skimming) implies the

complementary type of selection of participants to the other program (bottom scraping). In fact, this may not be a valid assumption.

3. An improper exclusion restriction can lead to extremely misleading inferences. The use of an alternative method of investigating selectivity bias based on the distributions of residuals provides a useful test of the validity of exclusion restrictions.

FOOTNOTES

1. Due to software pecularities, CHK estimated the two population selectivity bias model with a method slightly different from the method described in the text. However, the two methods produce results that differ only in sign.

2. CHK did use the weights in their original analysis, which employed least squares methods without explicit modelling of the school selection process.

3. We multiplied the design weights of the public school students by 988/125 to account for our sampling of only 125 of the 988 public schools in the HSB database.

4. Another difference between the weighted and unweighted data is the relative weight given to Catholic school students. As reported in Appendix A1, 44 percent of the students in the unweighted sample are in Catholic schools; in the weighted sample, only 7 percent are in Catholic schools. However, this does not play a major role in explaining the conflict between CHK's and Noell's results because the differences between the structure of the model for Catholic school students and public school students of the same ethnic group are minor relative to the differences in the structure of the model between ethnic groups.

APPENDIX A1

Means and Standard Deviations of Variables Used

	Unweighted Pooled (N=5500)	Weighted Pooled (5500)	Unweighted		
			White (3992)	Black (643)	Hispanic (865)
YBMATHRT	19.36 (7.13)	19.10 (7.22)	20.64 (7.05)	15.00 (5.60)	16.69 (6.47)
YBREADRT	9.48 (3.81)	9.26 (3.83)	10.03 (3.76)	7.73 (3.37)	8.23 (3.63)
SCHOOL	.44	.07 (.25)	.42	.50	.51
BLACK	.12	.10 (.30)			
HISPANIC	.16	.08 (.28)			
PARENTS	.75	.73 (.45)	.79	.49	.75
FEMALE	.54	.52 (.50)	.53	.57	.58
BBSESRAW	.03 (.74)	-.03 (.71)	.15 (.70)	-.17 (.71)	-.35 (.75)
NEAST	.25	.17 (.38)	.26	.25	.18
NCENT	.28	.27 (.45)	.32	.26	.09
SOUTH	.29	.36 (.48)	.25	.33	.41
CATHREL	.58	.34 (.47)	.56	.30	.85

Unweighted standard deviations of 0-1 variables are not given.

YBMATHRT = sum of correct answers on two math tests (38 items);
YBREADRT = number of correct answers on reading test (20 items);
BBSESRAW = composite SES measure constructed by the HSB sample designers.

The other 9 variables are 0-1 variables which take the value 1 according to these definitions: SCHOOL = Catholic; BLACK = black but not hispanic; HISPANIC = hispanic descent; PARENTS = both parents live at home; FEMALE = female; NEAST = resident in North East; NCENT = resident in North Central; SOUTH = resident in South; CATHREL = Catholic.

BLACK and HISPANIC are defined following CHK (1981a, p. 39, footnote to Table 3.1.1).

The variables BLACK...SOUTH constituted X_1.

The variables BLACK...CATHREL constituted X_2.

APPENDIX A2

Standard Errors for the Parameters of the Selectivity Bias Models

In this appendix it will be convenient in places to use vectors and matrices, rather than the subscripted scalars of Sections III and IV. To this end let $y = (y_1, \dots y_n)'$, and let the other variables be described by appropriate n-vectors, except $X_1 = (X_{11}, \dots X_{1n})'$, an (nxk) matrix, and $X_2 = (X_{21}, \dots X_{2n})'$, an (nxm) matrix.

(a) Selectivity bias and the probit model

u_{2i} is assumed to be standard normal and $E(u_{1i}/u_{2i}) = \rho\sigma_1 u_{2i}$. The arguments of this section parallel those of Heckman (1979) and Greene (1981).

$$\mathrm{var}(u_{2i}/\theta_i, z_i) = 1 + \theta_i h_i - h_i^2 ,$$

so that

$$\mathrm{var}(u_{1i}/\theta_i, z_i) = \sigma_1^2((1 - \rho^2) + \rho^2(1 + \theta_i h_i - h_i^2)).$$

Combining this information with equation (3.4), we can rewrite equation (3.1) in vector form:

$$y = X_1\beta_1 + \alpha z + ch + v, \qquad \text{(A1)}$$

where

$$E(v/\theta, z) = 0 , \quad E(vv'/\theta, z) = M , \quad c = \rho\sigma_1 .$$

0 is the n-vector of zeroes, and M is an (nxn) diagonal matrix whose (i,i)th entry is

$$M_{ii} = \sigma_1^2 + c^2(\theta_i h_i - h_i^2).$$

Note that the error vector v does not have homoscedastic elements.

Since β_2 is unknown, we must replace h in equation (A1) with $\hat{h}$, its estimate in Section III(B). Equation (A1) therefore becomes,

$$y = A\gamma + \eta . \qquad \text{(A2)}$$

where,

$A = (X_1, z, \hat{h})$, an n x (k+2) matrix;

$\gamma = (\beta_1 ; \alpha, c)'$, a (k+2)-vector;

$\eta = c(h - \hat{h}) + v$.

Performing ordinary least squares on equation (A2) yields consistent estimates, $\hat{\gamma}$, of γ .

Under general conditions on the elements of z and X_2 (Amemiya, 1973; Jennrich, 1969),

$n^{\frac{1}{2}}(\beta_2 - \beta_2) \rightarrow N(0,\Omega)$ as $n \rightarrow \infty$ for some positive definite (mxm) matrix Ω , and

$n^{\frac{1}{2}}(\hat{h} - h) \rightarrow N(0,\Delta X_2 \Omega X_2'\Delta)$ as $n \rightarrow \infty$, conditional on z and θ , where $\Delta = \partial h/\partial\theta$ is an (nxn) diagonal matrix such that,

$$\Delta_{ii} = \partial h_i/\partial\theta_i = h_i^2 - \theta_i h_i .$$

To assign appropriate standard errors to the estimates of α , β_1 , and c , we need to determine the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\gamma} - \gamma)$.

$$n^{\frac{1}{2}}(\hat{\gamma} - \gamma) = n(A'A)^{-1} . \; n^{-\frac{1}{2}}A'[c(h - \hat{h}) + v] .$$

Now,

$$\operatorname*{plim}_{n\to\infty} n(A'A)^{-1} = \operatorname*{plim}_{n\to\infty} n\,(z\;X_1\;h)'(z\;X_1\;h)^{-1} = B$$

, a positive definite symmetric (k+2) x (k+2) matrix, if the Amemiya-Jennrich conditions hold.

Then, $n^{\frac{1}{2}}(\hat{\gamma} - \gamma) \rightarrow N(0,B\psi B)$ as $n \rightarrow \infty$,

where

$$\psi = \operatorname*{plim}_{n\to\infty} (\psi_1 + \psi_2) .$$

$$\psi_1 = n^{-1} A'MA \ ,$$

$$\psi_2 = c^2 n^{-2} (A'\Delta X_2)\Omega(A'\Delta X_2)' \ .$$

An estimate for the variance-covariance matrix of $\hat{\gamma}$ is thus

$$(A'A)^{-1}[A'\hat{M}A + \hat{c}^2(A'\hat{\Delta}X_2)\hat{\Sigma}(X_2{}'\hat{\Delta}A)](A'A)^{-1} \qquad (*)$$

Here $\hat{\Sigma}$ is the estimated variance-covariance matrix of $\hat{\beta}_2$ from the probit $(\lim\limits_{n\to\infty} n\hat{\Sigma} = \Omega)$, $\hat{c}$ is the estimated coefficient on $\hat{h}$ from regression equation (A2),

$$\hat{\Delta}_{ii} = \hat{h}_i{}^2 - \hat{\theta}_i\hat{h}_i \ ,$$

$$\hat{M}_{ii} = \hat{\sigma}_1{}^2 - \hat{c}^2\hat{\Delta}_{ii} \ ,$$

$$\hat{\sigma}_1{}^2 = n^{-1}\sum_{i=1}^{n}\hat{\eta}_i{}^2 + \hat{c}^2 n^{-1}\sum_{i=1}^{n}\hat{\Delta}_{ii} \ .$$

$\hat{\eta}$ is the vector of observed residuals from regression equation (A2). The correlation between the achievement and selection errors $(u_1$ and $u_2)$ can be estimated by $\hat{\rho} = \hat{c}/\hat{\sigma}_1$. $\hat{c}$, $\hat{M}$, $\hat{\Delta}$, $\hat{\rho}$, and $\hat{\sigma}_1{}^2$ are consistent estimators of c , M , Δ , ρ , and $\sigma_1{}^2$.

Heckman's censored sample model can be estimated with minor modifications of the methodology presented here. In that model, one only observes y for those individuals in the sample for whom $u_{2i} > \theta_i$. That is, z_i is 1 if y_i is observed. Suppose the first n_1 individuals have observations on y , and the last $n - n_1$ do not. First perform maximum likelihood probit for all n cases, then construct $\hat{h}_i = f(\hat{\theta}_i)/(1 - F(\hat{\theta}_i))$ only for the first n_1 cases, and estimate eq.(A2) for these n_1 individuals, setting $X_1 = (X_{11}, \ldots X_{1n_1})'$, $A = (X_1, \hat{h})$, and $\gamma = (\beta_1, c)'$. Finally, construct $(*)$ with $X_2 = (X_{21}, \ldots X_{2n_1})'$ and n replaced by n_1 in the estimation of M , Δ , and $\sigma_1{}^2$.

(b) Selectivity bias controlled by the linear probability model

We suppose that u_{2i} is uniform on $[0,1]$ for each $i=1,\ldots n$, rather than being standard normal. So $E(u_{2i}) = 1/2$, and $\text{var}(u_{2i}) = 1/12$. We also assume that $E(u_{1i}/u_{2i}) = \rho\sigma_1\sqrt{12}\,(u_{2i} - 1/2)$ (Olsen, 1980).

Then,

$$P[z_i=1] = X_{2i}'\tilde{\beta}_2 \tag{A3}$$

Here $X_{2i}'\tilde{\beta}_2 = 1 - \theta_i$, where $\theta_i = -X_{2i}'\beta_2$, as defined in Section III(A).

$$E(u_{2i}/\theta_i,z_i) = (\theta_i + z_i)/2\ ,$$

$$\text{var}(u_{2i}/\theta_i,z_i) = (\theta_i - z_i)^2/12\ .$$

Let $\delta = \rho\sigma_1\sqrt{3}$. Then,

$$E(u_{1i}/\theta_i,z_i) = \delta(\theta_i + z_i - 1)\ ,$$

$$\text{var}(u_{1i}/\theta_i,z_i) = \sigma_1^2 - (\delta^2/3)(1 - (\theta_i - z_i))^2 = M_{ii}\ \text{, say.}$$

Equation (3.1) therefore can be rewritten, in vector form:

$$y = X_1\beta_1 + \alpha z + \delta s + v \tag{A4}$$

where now

$s_i = z_i - (1 - \theta_i)$ (1 is the n-vector of 1's here.)

$E(v/\theta,z) = 0$; (0 is the n-vector of 0's here.)

$E(vv'/\theta,z) = M$, a diagonal (nxn) matrix whose (i,i)th entry is M_{ii}.

It can be seen that s, the auxiliary regressor, has ith component which is the difference between the actual value of z_i and the probability that z_i is 1. To avoid collinearity between X_1, z, and s in equation (A4), we must identify at least one variable that is in X_2, but not in X_1; that is, a variable that affects school choice, but does not directly affect ability.

Use the fitted values, $\hat{z}$, from the regression of z on X_2 to construct $\hat{s} = z - \hat{z}$. Then,

$$y = A\gamma + \nu \qquad \text{(A5)}$$

In this specification,

$A = (X_1, z, \hat{s})$, an n x (k+2) matrix;

$\gamma = (\beta_1, \alpha, \delta)'$, a (k+2)-vector;

$\nu = \delta(s - \hat{s}) + v$.

If the usual conditions hold on X_2 and z (Theil, 1973),

$\operatorname*{plim}_{n\to\infty} n(A'A)^{-1} = \operatorname*{plim}_{n\to\infty} n[(X_1\ z\ s)'(X_1\ z\ s)]^{-1} = B$, a positive definite symmetric (k+2) x (k+2) matrix, and

$n^{\frac{1}{2}}(\hat{\beta}_2 - \beta_2) \to N(0\ ,\ \Omega)$ as $n \to \infty$, for some positive definite (mxm) matrix Ω, which need not be the same as the Ω of part (a) above.

Thus, $n^{\frac{1}{2}}(\hat{s} - s) \to N(0\ ,\ X_2\Omega X_2')$ as $n \to \infty$, and we can find the limiting distribution of $n^{\frac{1}{2}}(\hat{\gamma} - \gamma)$ in a manner similar to before.

$n^{\frac{1}{2}}(\hat{\gamma} - \gamma) \to N(0\ ,\ B\Psi B)$ as $n \to \infty$,

where $\psi = \operatorname*{plim}_{n\to\infty} (\psi_1 + \psi_2)$.

$\psi_1 = n^{-1}A'MA$,

$\psi_2 = \delta^2 n^{-2}(A'X_2)\Omega(A'X_2)'$.

An estimate for the variance-covariance matrix of $\hat{\gamma}$ is then

$$(A'A)^{-1}[A'\hat{M}A + \hat{\delta}^2(A'X_2)\hat{\Sigma}(X_2'A)](A'A)^{-1} \qquad (\dagger)$$

Here $\hat{\Sigma}$ is the estimated variance-covariance matrix of the coefficients from the linear probability model $(\lim_{n\to\infty} n\hat{\Sigma} = \Omega)$, $\hat{\delta}$ is the estimated coefficient on $\hat{s}$ from regression equation (A5),

$$\hat{M}_{ii} = \sigma_1^2 - \hat{q}_i ,$$

$$\hat{\sigma}_1^2 = n^{-1}\sum_{i=1}^{n}\hat{\nu}_i^2 + n^{-1}\sum_{i=1}^{n}\hat{q}_i ,$$

$$\hat{q}_i = (\hat{\delta}^2/3)(1 - (\hat{\theta}_i - z_i)^2) = (\hat{\delta}^2/3)(1 - (\hat{z}_i - (1 - z_i))^2)$$

$$\hat{\rho} = \hat{\delta}/\sigma_1\sqrt{3} .$$

$\hat{\nu}$ is the vector of observed residuals from regression equation (A5). $\hat{\Delta}$, $\hat{M}$, $\hat{\rho}$, and $\hat{\sigma}_1^2$ are consistent estimators of the true parameter values.

The censored sample model of Olsen can be estimated by running the linear probability model on all n observations, then using least squares on equation (A5) for the n_1 individuals for whom y is observed, setting

$\hat{s} = 1 - \hat{z}$, $X_1 = (X_{11},\ldots X_{1n_1})'$, $A = (X_1, \hat{s})$, and

$\gamma = (\beta_1, \delta)'$, Finally, (†) is constructed with

$X_2 = (X_{21},\ldots X_{2n_1})'$, and n replaced by n_1 in the estimation of

σ_1^2 , q , and M.

It is a fairly straightforward matter to compute the standard errors given by the diagonal elements of (*) or (†) once the achievement regression (equation (A2) or (A5)) has been run. The matrix $(A'A)^{-1}$ is just the estimated variance-covariance matrix of the coefficients of this regression, multiplied by the estimated variance of the regression; $\hat{\Sigma}$ is the estimated variance-covariance matrix from the probit or least squares regression of z on X_2 , and the other estimates are easily constructed.

One problem in finite samples is that there is no guarantee that σ_1^2 will be positive in either part (a) or (b). This can result in negative diagonal elements in (*) or (†) (Greene, 1981).

Once we condition on X_2 , the ith residual from the linear probability model has variance $\theta_i(1 - \theta_i)$. One way to induce homoscedastic errors is to run the linear probability model using ordinary least squares, then run weighted least squares on the same equation, giving the ith observation a weight $[\hat{z}_i(1-\hat{z}_i)]^{-\frac{1}{2}}$. Such a scheme requires each predicted probability, $\hat{z}_i$, to lie between 0 and 1.

APPENDIX A3

Two Stage Least Squares and the s Method

Recall that $\hat{s} = z - \hat{z}$, where $\hat{z}$ is the vector of fitted values from regression of z on X_2. Ordinary least squares on the achievement equation,

$$y = X_1\beta_1 + \alpha z + \delta\hat{s} + \nu$$

yields a vector of predicted scores, $\hat{y}$:

$$\hat{y} = X_1\hat{\beta}_1 + \hat{\alpha}z + \hat{\delta}\hat{s} ,$$

where $\hat{\beta}_1$, $\hat{\alpha}$, and $\hat{\delta}$ are the least squares estimates of β_1 , α , and δ .

So,

$$\begin{aligned} \hat{y} &= X_1\hat{\beta}_1 + \hat{\alpha}z + \hat{\delta}(z - \hat{z}) \\ &= X_1\hat{\beta}_1 + \hat{\alpha}(\hat{z} + (z - \hat{z})) + \hat{\delta}(z - \hat{z}) \\ &= X_1\hat{\beta}_1 + \hat{\alpha}\hat{z} + (\hat{\alpha} + \hat{\delta})(z - \hat{z}) \end{aligned}$$

But, if we use ordinary least squares to regress z on X_2 , then the residual vector, $z - \hat{z}$, is orthogonal to $\hat{z}$ and X_2 , hence also to X_1 , since X_2 contains all the variables in X_1 . Therefore, the estimates of β_1 and α from the s method will be the same as those derived by two stage least squares on the model,

$$y = X_1\beta_1 + \alpha z + u_1$$

$$z = X_2\beta_2 + u_2$$

Moreover, if we use two stage least squares, then the standard errors of the estimated coefficients will be correct; however, two stage least squares does not permit an immediate test of selectivity bias.

REFERENCES

Amemiya, T. "Regression Analysis When the Dependent Variable is Truncated Normal." Econometrica 41 (1973), 997-1017.

Barnow, S.N., G.G. Cain, and A.S. Goldberger. "Issues in the Analysis of Selectivity Bias," in Evaluation Studies, Vol. 5 (1980), E.W. Stromsdorfer and G. Farkas (eds.). Beverly Hills, CA: Sage Publications.

Coleman, J., S.N. Hoffer, and S. Kilgore. Public and Private Schools. Report to the National Center for Education Statistics by the National Opinion Research Center (Nov. 1981a).

Coleman, J., S.N. Hoffer, and S. Kilgore. "Questions and Answers: Our Response." Harvard Educational Review 51 (1981b), 526-545.

Farkas, G., R.J. Olsen, and E.W. Stromsdorfer. "Reduced-Form and Structural Models in the Evaluation of the Youth Entitlement Demonstration," in Evaluation Studies, Vol. 5 (1980), E.W. Stromsdorfer and G. Farkas (eds.). Beverly Hills, CA: Sage Publications.

Goldberger, A. S. "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations." Discussion Paper 123-172, University of Wisconsin, Institute for Research on Poverty (1972).

Goldberger, A. S. "Abnormal Selection Bias." Discussion Paper 8006, University of Wisconsin, Social Systems Research Institute (1980).

Goldberger, A. S. and Cain, G.G. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report." Sociology of Education 55 (April/July 1982).

Greene, W.H. "Sample Selection Bias as a Specification Error: Comment." Econometrica 49 (May 1981).

Gronau, R. "The Effect of Children on the Housewife's Value cf Time." Journal of Political Economy 81 (1973).

Gronau, R. "Wage Comparisons--A Selectivity Bias." Journal of Political Economy 82 (1974), 1119-1144.

Heckman, J.J. "Shadow Prices, Market Wages, and Labor Supply." Econometrica 42 (1974), 679-694.

Heckman, J.J. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." Annals of Economic and Social Measurement 5 (1976), 475-492.

Heckman, J.J. "Dummy Endogenous Variables in a Simultaneous Equations System." Econometrica 46 (1978), 931-960.

Heckman, J.J. "Sample Bias as a Specification Error." Econometrica 47 (1979), 153-162.

Jennrich, R. "Asymptotic Properties of Nonlinear Least Squares Estimators." Annals of Mathematical Statistics 40 (1969), 633-643.

McGuire, T.G. Financing Psychotherapy: Costs, Effects, and Public Policy. Cambridge, Mass.: Ballinger, 1981.

Maddala, G.S. and L. Lee. "Recursive Models with Qualitative Endogenous Variables." Annals of Economic and Social Meausrement 5 (1976), 525-545.

Mallar, C.D., S.H. Kerachsky, and C.V.D. Thornton. "The Short-Term Economic Impact of the Job Corps Program," in Evaluation Studies, Vol. 5 (1980), E.W. Stromsdorfer and G. Farkas (eds.). Beverly Hills, CA: Sage Publications.

Noell, J. "Public and Catholic Schools: A Reanalysis of 'Public and Private Schools'." U.S. Department of Education (manuscript) (1981).

Noell, J. "Public and Catholic Schools: A Reanalysis of 'Public and Private Schools'." Sociology of Education 55 (April/July 1982).

Olsen, R.J. "A Least Squares Correction for Selectivity Bias." Econometrica 48 (Nov. 1980), 1815-1820.

Olsen, R.J. "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator." International Economic Review 23 (Feb. 1982).

Orazem, P. "Black-White Differences in Human Capital Investment and Earnings." PhD Dissertation, Yale University, forthcoming.

Theil, H. Principles of Econometrics. New York: Wiley & Sons, 1971.

Willis, R.J. and S. Rosen. "Education and Self-Selection." Journal of Political Economy S7-S36 (Oct. 1979).

Selection and Survival in the Teacher Labor Market

by

Richard J. Murnane

Institution for Social and Policy Studies
Yale University

July 16, 1982

I would like to thank Charles Lindblom, Richard Nelson, Sharon Oster, and Edward Pauly for helpful comments on earlier drafts, Randall Olsen and Jon Peck for advice on statistical issues, and Norman Futami for first-rate research assistance. The research on which this paper is based was supported by grants from the Spencer Foundation and the National Institute of Education.

I. INTRODUCTION

Fears about deterioration of the quality of U.S. public education, particularly in urban areas, have led to a concern about changes in the quality of the stock of public school teachers. One determinant of the quality of the teaching stock, the quality of new entrants, has been investigated extensively (Weaver, 1978) and the evidence is that the average quality of college graduates entering the teaching profession, as measured by scholastic aptitude test scores, has declined in recent years.

The second determinant of the quality of the teaching stock, the pattern of turnover among teachers of differing quality, has not been investigated thoroughly, primarily due to a lack of data on the performance of individual teachers.[1/] Consequently, it is not known whether the pattern of teacher turnover (quits and involuntary terminations) contributes to the alleged decline in the quality of the teaching stock or alleviates the problem caused by the low average quality of recent entrants. This paper addresses this question by examining whether highly productive teachers remain teaching in a large urban school district longer or less long than less productive teachers do.

Three measures of teaching performance are used in this analysis, all of which provide more direct information about productivity than do the demographic characteristics and aptitude test scores that have been used as proxies for teaching quality in past research on teacher labor markets. Examination of the sensitivity of the empirical results to the choice of instrument used to measure productivity sheds additional light on how teacher labor markets work.

II. WHY WOULD TURNOVER BE SYSTEMATICALLY RELATED TO PRODUCTIVITY?

To provide a basis for interpreting the empirical results reported later in the paper, it is instructive to consider alternative models that predict a systematic relationship between teacher turnover and productivity. If employers had complete information about the productivity of all applicants before hiring decisions were made, and if, at that time, applicants had complete information about the streams of compensation and job satisfaction for possible jobs, then only mutually advantageous employment agreements (job matches) would be made. Under these circumstances, turnover would be unrelated to worker productivity or to the characteristics of jobs. Thus, models that predict selective attrition of the least or most productive workers assume that information becomes available after a job match is made concerning either the quality of the existing match or the quality of an alternative job match (Jovanovic, 1979). I focus on information regarding the existing match.

One set of assumptions that leads to the prediction that low productivity teachers will remain less long in a public school district than high productivity teachers will is that administrators acquire reliable information about productivity only by observing teachers on the job and that administrators act on this newly acquired information to alter the terms of employment of individual teachers. Most observers agree that administrators do acquire important information about productivity by observing teachers at work and by listening to feedback (e.g., from parents). The remaining question is whether they act on this information. Since the contracts of most public school teachers specify that salaries are determined exclusively by seniority and degrees, performance-based wage adjustments, which are hypothesized to be important in the private sector (Jovanovic, 1979), are not a possible course of action for public school administrators. Alternative responses that are possible include the termination and threat of termination of the contracts of teachers perceived to be unproductive. Formal terminations are rare in most school districts. However, resignations that result from threats of contract termination may be common.

A second set of assumptions, complementary to the first and leading to the same prediction of selective attrition of the least productive teachers, concerns applicants' predictions of their own productivity and job satisfaction. Although most applicants for public school teaching positions participate in practice teaching as part of their training, this is typically done under the supervision of an experienced teacher who protects the trainee from the discipline problems that often accompany poor teaching performance and are the greatest source of low job satisfaction. Consequently, practice teaching may provide little reliable information about a trainee's potential productivity as an independent teacher. Teachers who find that they are not successful in running their own classroom may resign in response to the low job satisfaction that stems from poor teaching performance and the accompanying discipline problems.

Among the sets of assumptions about information flows that lead to the opposite prediction--selective attrition of the most productive teachers--perhaps the most compelling concerns the deterioration in the financial status of the teaching profession during the 1970's. On the basis of salary trends in the 1960's, teachers hired in 1970 could expect continued improvement in the financial status of their profession, both absolutely and relative to alternative occupations. However, the 1970's proved to be a disappointing period for teachers as inflation coupled with fiscal problems in many cities led to an erosion of the profession's earlier financial gains. If teachers felt that their productivity in the classroom predicted their ability to succeed in an alternative occupation, then the most productive teachers would be the most likely to respond to the unexpected salary disappointments by resigning.

III. DATA

The data for this study include information on a sample of elementary school teachers who worked in the early 1970's in an urban school system, known here as Eastcity. Thirty-two percent of the teachers are black; 89 percent are female. All of the teachers taught either a second or third grade class in an inner city school. Each of the teachers worked with a single group of students for the entire school day.

Data were also collected describing 1,545 black students taught by the teachers in the sample. The data include family background information and reading and mathematics test scores for each student. Each student was tested twice: at the end of the school year before the student entered the sampled teacher's classroom, and also at the end of the school year in which the student studied with the teacher in the sample. The availability of longitudinal information on students' reading and mathematics skill levels permitted analysis of the value added of an additional year in school spent with a given teacher. These data were originally collected for a study investigating the impact of teachers on students' cognitive skills (Murnane, 1975).

Two estimates of the productivity of each teacher were derived by estimating equation (1): first, using students' reading scores as the dependent variable, then using students' mathematics scores.[2/]

$$A_{in} = a_1 A_{i,n-1} + \sum_k d_k D_{ik} + \sum_{j=1}^{104} t_j T_{ij} + e_i \qquad (1)$$

where

A_{in} = the ith student's score on a test of reading (math) skills at the end of the school year, n

$A_{i,n-1}$ = the ith student's score on a test of reading (math) skills at the end of the previous school year, n-1

D_{ik} = the kth demographic characteristic describing the ith child and his/her family[3/]

T_{ij} = 1, if the ith child was in the classroom of the jth teacher in school year,n

= 0, otherwise

t_j = the estimate of the jth teacher's value added in teaching reading (math) skills to his/her students

e_i = a student specific error term with mean zero, and constant variance.

F tests were used to test the null hypothesis that $t_j = t_m$, $j,m = 1,,,104$, $j \neq m$. This test can be interpreted as asking the following question: If one wants to predict a child's achievement at the end of a school year, and one already knows the child's achievement level at the end of the previous school year along with some characteristics of the child's background, is it also important to know in which classroom the child spent the school year? The results indicate that there are significant differences in the average amount of reading and mathematics achievement that took place among students in different classrooms in the sample, after accounting for students' backgrounds and prior skill levels.[4/] Not only are these differences statistically significant, but they are also sufficiently large to be of policy interest. Expressed in terms of grade equivalents, the average gains for students in particular classrooms ranged from four months to 12 months. The range of variation in mathematics achievement gains was even larger.

In this study, which follows the approach of Hanushek (1972), the estimated average achievement test gains of students in reading and mathematics (the values of t_j in equation (1)) are interpreted as two measures of the productivity of classroom teachers. It is important to note that, according to the principals of the schools from which the data were drawn, students were not grouped by ability in their assignment to teachers. This is a necessary condition for interpreting the t_i coefficients as measures of the value added of individual classroom teachers. Data from two other schools, in which students were grouped by ability, were excluded from the analysis.

A third measure of teacher productivity was provided by supervisors' evaluations. In this school system every teacher is evaluated periodically by the school principal using a standardized form that includes 19 dimensions of performance. The variable used to measure performance in this study is the sum of these ratings, standardized to have a mean of zero and a standard deviation of one.[5/]

Each of the three measures of teacher performance has a desirable attribute. The measure of effectiveness in teaching reaching is valuable because of the consensus that teaching children to read should be a high priority objective of school programs. The strength of the measure of effectiveness in teaching mathermatics is that prior research indicates that children's mathematics skills are more sensitive to school instruction and less sensitive to family background than are reading skills (Murnane, 1975). Consequently, the measure of effectiveness in teaching mathematics may be less contaminated by unmeasured differences in student backgrounds than is the measure of performance in teaching reading.

The principal's evaluation has the potential advantage of capturing dimensions of a teacher's performance not reflected in student test score gains.[6/] A similar argument has been made by the authors of several recent studies that have used supervisors' performance ratings in analyzing the career patterns of workers in the private sector (Medoff and Abraham, 1980, 1981).

Statistics describing the distributions of the three performance measures and the correlations among them are presented in Table 1. (This information is presented for the full sample, for subsamples defined by the number of years a teacher had been in the school system when performance was assessed, and for the subsamples used in the empirical work described below.) The Pearson simple correlations indicate that, while the performance measures are positively correlated for the full sample and for most subsamples, the correlations are quite low. This is interesting in itself because it demonstrates that if terms of employment were a function of assessed performance, the terms offered to any individual teacher would

---

Table 1 Here

Table 1

Means and Standard Deviations of Variables Describing Teachers and Their Performance

	All teachers	Teachers grouped by the number of years in the school system when performance was assessed						Samples used in estimation[a]		
								Sample 1	Sample 2	Sample 3
		1	2	3	4	5 to 10	GE 11			
	N=104	N=37	N=21	N=11	N=9	N=15	N=11	N=30	N=52	N=60
Teacher race, black	0.32	0.19	0.52	0.27	0.44	0.40	0.18	0.17	0.33	0.38
Teacher sex, female	0.89	0.92	0.90	0.82	0.67	0.93	1.00	0.90	0.92	0.85
Performance in teaching reading (Read)	1.53 (0.41)	1.37 (0.34)	1.65 (0.37)	1.76 (0.58)	1.56 (0.22)	1.68 (0.44)	1.40 (0.42)	1.36 (0.36)	1.49 (0.38)	1.56 (0.41)
Performance in teaching mathematics (Math)	1.01 (0.44)	0.83 (0.38)	1.11 (0.49)	1.10 (0.55)	1.17 (0.49)	1.23 (0.35)	0.95 (0.28)	0.82 (0.41)	0.93 (0.44)	1.00 (0.47)
Principal's evaluation of teaching performance (Eval)[b]	-0.01 (1.00) (N=96)	-0.46 (0.97) (N=29)	0.10 (1.10) (N=21)	0.38 (0.97) (N=10)	-0.23 (1.09) (N=7)	0.67 (0.38) (N=12)	0.34 (0.65) (N=10)	-0.59 (0.99) (N=29)	-0.20 (1.06) (N=52)	-0.06 (1.00) (N=57)
Zero order correlation between										
Read, Math	0.38	0.27	-0.04	0.61	0.31	0.45	-0.23	0.33	0.26	0.41
Read, Eval	0.22	0.11	0.03	0.35	-0.58	0.26	0.04	0.07	0.15	0.17
Math, Eval	0.44	0.44	0.42	0.50	-0.04	0.22	0.32	0.45	0.47	0.36

---

[a] Sample 1 includes only the 30 teachers who were in their first year of teaching when performance was assessed. Seven of the 37 teachers who were in their first year of teaching in the Eastcity school system when performance was assessed had taught previously in other school systems and were excluded from Sample 1. Five of these seven teachers did stay in the Eastcity system for at least two years and were included in Sample 2.

[b] For a small percentage of teachers, the principals' evaluations were missing from the personnel folders, and consequently the empirical work that uses the evaluations is based on slightly smaller samples.

be very sensitive to the choice of instrument used to measure performance.

In order to investigate the attrition of teachers from the school system, it was necessary to gather information on the number of years that each teacher remained in the school system after the date at which performance was assessed. This information, which was collected from the school system personnel records in 1980, indicates that most attrition among teachers in this sample took place in the first years of teaching. In fact, 88 percent of the teachers who had been in the system at least five years when performance was assessed in the early 1970's were still in the system in 1980 (compared to 27 percent of the teachers who were in their first year in the system when performance was assessed). Consequently, if significant selective attrition did occur in this sample, it occurred among teachers who were in their first years of teaching during the period of observation (1971-1980).[6]

IV. MODELS, SAMPLE DEFINITIONS, AND ESTIMATION TECHNIQUE

The first question examined in the empirical work is whether the length of stay of teachers in the Eastcity school system was systematically related to their productivity during their first year of teaching. This question was addressed by estimating equation (2) for the 30 teachers who had just started their careers when the performance assessments were made (referred to as sample 1):

$$Ln(Yrs_i) = a + p(P_{i1}) + u_i \quad (2)$$

where

$Ln(Yrs_i)$ = the natural logarithm of the total number of years that the ith teacher taught in the Eastcity school system

P_{i1} = one of the three estimates of the ith teacher's productivity during his or her first year of teaching

u_i = a teacher specific error term with mean zero and constant variance.

Given our interest in understanding the timing of any selective attrition that did occur, it was instructive to investigate two additional questions:

> Among teachers who stayed in the Eastcity school system for more than one year (more than three years), was length of stay after the first year (after the third year) systematically related to productivity?

The first of these additional questions permits examination of whether selective attrition occurred only during the first year of teaching, the time during which the greatest amount of new information about productivity was probably generated, or whether it also took place among teachers who stayed in the system for more than one year. Similarly, investigation of the second question permits examination of whether selective attrition occurred among tenured teachers. (The tenure decision in this school system is made during a teacher's third year in the system.)

To investigate these additional questions, two samples were formed. Sample 2 consists of teachers who were in their second year of teaching in the Eastcity system when their performance was assessed (N=21) plus the subset of teachers in their first year of teaching in the system when performance was assessed who stayed at least a total of two years in the system (N=31). To make the dependent variable, labelled $\mathrm{Ln}(\mathrm{Yrsafter}_{i1})$, have a consistent meaning for the two subsamples of sample 2, it was defined as the natural logarithm of the number of years the ith teacher remained in the system after the first year.

Similarly, sample 3 consists of teachers who were in their fourth year of teaching when their performance was assessed (N=9) plus those teachers who were in their first, second, or third year in the system when performance was assessed and who stayed in the system at least a total of four years (N=24, N=17, N=10, respectively). For this sample, the dependent variable, $\mathrm{Ln}(\mathrm{Yrsafter}_{i3})$, was defined as the natural logarithm of the number of years the ith teacher taught in the system after the third year, the year in which tenure was awarded.

Equation (3) is the model we would like to have estimated to address the questions dealing with the timing of selective attrition.

$$\mathrm{Ln}(\mathrm{Yrsafter}_{ij}) = a + p(P_{i1}) + u_i \qquad j=1,3 \tag{3}$$

However, this was not possible because P_{i1} was not observed for all of the teachers in samples 2 and 3. Some of the teachers in these samples had had more than one year of experience at the time their performance was assessed. Prior research (Murnane, 1975) indicated that the productivity of teachers increases over their first three years of teaching, with the greatest gains made during the first year on the job. Consequently, to make the performance measures comparable among teachers who had different amounts of experience at the time their performance was assessed, it was necessary to model explicitly the relationship between experience and productivity.

Following Murnane (1975) the following specification was adopted:

$$P_{in} = P_{i1} + b[f(n)] + v_i \qquad (4)$$

where

n = the number of years a teacher had taught as of the end of the year in which performance was assessed

P_{in} = the productivity of the ith teacher during his or her nth year of teaching, as estimated by one of the three performance measures

$f(n) = 0$ if $n=1$

$= 2$ if $n=2$

$=3$ if $n \geq 3$

v_i = a teacher specific error term with mean zero and constant variance.

Substituting equation (4) into equation (3) produces equation (5) which was estimated to investigate whether selective attrition occurred among teachers who had already survived in the Eastcity system for two (or four) years.

$$\mathrm{Ln}(\mathrm{Yrsafter}_{ij}) = a + p(P_{in}) + d[f(t)] + w_i \qquad j=1,3 \qquad (5)$$

where

$d = -(p)(b)$

$w_i = u_i - p(v_i)$

Sensitivity analysis indicated that the estimates of the coefficients of equation (5) were not substantively changed when alternative plausible specifications of $f(n)$ were adopted.

Maximum likelihood tobit analysis was used to estimate equation (2) for sample 1 and equation (5) for samples 2 and 3, to take account of the presence in these samples of teachers who were still teaching in the

Eastcity system when the personnel records were examined in 1980. For these teachers, the observed length of stay understates their true length of stay. With such a censored dependent variable, tobit analysis provides consistent estimates of the parameters of these equations, while ordinary least squares does not. The results of all of the estimations are reported in Table 2.

Table 2 Here

V. RESULTS

A. Productivity measured by student test score gains in reading and mathematics

There is no statistically significant evidence of selective attrition of the least productive or most productive teachers, when productivity is measured by student test score gains. Although the coefficients on the productivity measures are consistently positive, suggesting that the most productive teachers may remain in the Eastcity system longer than the least productive teachers do, the consistently low values of the t statistics indicate that the null hypothesis that length of stay is unrelated to productivity cannot be rejected.

This conclusion of no selective attrition is also supported by the summary statistics in Table 1. These cross-sectional data show that the average performance in reading and mathematics instruction by teachers who had five or more years of experience in the school system at the time performance was assessed was slightly higher than the average performance of first year teachers. This is consistent with the conclusion of no

Table 2

Tobit estimates used to test hypotheses concerning selective attrition

(t statistics in parentheses)

Dependent Variable	Sample 1 (eq.2) N=30 ln(length of stay in school system)				Sample 2 (eq.5) N=52 ln(length of stay in school system after first year)				Sample 3 (eq.5) N=60 ln(length of stay in school system after third year)			
erformance in eaching reading	0.74 (1.30)				0.39 (0.95)				0.18 (0.46)			
erformance in eaching math		0.27 (0.54)				0.40 (1.18)				0.22 (0.65)		
rincipal's valuation			0.54* (2.98)				0.17 (1.20)				0.00 (0.01)	
rincipal's valuation-- eacher is white				0.55* (2.69)				0.42* (2.56)				0.10 (0.44)
rincipal's valuation-- eacher is black				0.44 (1.16)				-0.12† (-0.62)				0.07 (0.32)
eacher is black				0.42 (0.79)				1.03* (3.39)				0.82* (2.49)
(N)					-0.05 (-0.45)	-0.05 (-0.44)	-0.07 (-0.57)	-0.30* (-2.43)	0.07 (0.56)	0.07 (0.60)	0.06 (0.54)	-0.02 (-0.17)
ntercept	0.71 (0.91)	1.48* (3.20)	2.07* (9.63)	2.00* (8.47)	1.12 (1.88)	1.32* (3.81)	1.75* (7.94)	1.72* (8.48)	1.38* (2.34)	1.65* (3.91)	1.82* (8.22)	1.41* (5.44)
tandard error of he regression	4.94	4.94	4.91	5.12	4.73	4.72	4.72	4.81	4.35	4.35	4.32	4.40
roportion of sample ith limiting value f dependent variable	30				29				42			

* Significantly different from zero at .05 level on two-tailed test.

† Significantly different at .05 level from coefficient on principal's evaluation for white teachers.

selective attrition since we know that first year teachers become more productive as they gain experience (Murnane, 1975). The data in Table 1 also show that the variation in performance in mathematics instruction among teachers with five or more years of experience was almost as large as the variation in performance among first year teachers. The variation in performance in reading instruction was actually larger among senior teachers than among first year teachers. Thus, the sample contained some very productive senior teachers and some senior teachers who were much less successful in helping their students to improve reading skills.

By themselves, these results are consistent either with the assumption that the student test score gains do not provide valuable new information about teacher productivity that enables administrators and/or teachers to reevaluate the quality of the job match, or that the test score gains do provide new information, but that neither administrators nor teachers act on this information in a way that leads to selective attrition. The results relating length of stay to the principals' evaluations indicate that the second of these assumptions is not correct for untenured teachers.

B. Productivity measured by principals' evaluations

Teachers who receive low performance ratings from their school principals in their first years of teaching are more likely to leave the school system than are teachers who receive higher performance ratings. This relationship is most clear-cut in sample 1, for which the dependent variable measures length of stay from the beginning of teaching in the system. When length of stay is measured from the beginning of the second year in the system (sample 2), evidence of continued selective attrition is present although the positive relationship between teaching performance and length of stay is only statistically significant for white teachers. Among teachers who remained in the system for at least four years (sample 3) and consequently had achieved tenured status, there is no evidence of continued selective attrition.

The distribution of performance ratings for the cross-section of teachers observed in the early 1970's is consistent with the results of the longitudinal analysis of teachers' mobility decisions. As shown in Table 1, the average performance rating of teachers with five or more years of experience is higher than the average rating of first year teachers. Also, the standard deviation of the ratings of senior teachers is only half the size of the standard deviation of the ratings of first year teachers. One would expect these patterns to characterize the distribution of performance ratings in the stock of teachers at one point in time, given that teachers who do not receive satisfactory performance ratings in their first years of teaching leave the school system.

It is interesting that the evidence on selective attrition is sensitive to the instrument used to measure productivity. The evidence suggests that new information is generated during the first years on the job, and this information does influence length of stay. However, the important information is reflected in principals' evaluations, not in student test score gains. One possible explanation is that the student test score gains reflect the effects of factors beyond the control of teachers, such as peer group influences, more than they reflect teachers' abilities and efforts. Thus, the evaluations may provide more reliable information about productivity than the student test score gains do and this is why only the evaluations are systematically related to length of stay.

It is also possible that the evaluations reflect administrators' preferences for particular types of teachers or teaching styles, preferences that are only weakly correlated with teaching performance as measured by students' progress. It is not possible to differentiate between these explanations. However, it is important to keep in mind that irrespective of whether the evaluations provide reliable information about productivity or not, they do matter, in that teachers who receive low evaluations in their first years of teaching are particularly likely to leave the school system.

One curious finding is that in sample 2, the performance ratings of white teachers are significantly related to their length of stay in the school system, but this is not true for black teachers. I have no compelling explanation for this difference. However, it does suggest that it is worthwhile exploring in other data sets that include performance ratings whether relationships between performance ratings and career paths are different for black and white workers.

VI. SUMMARY

The findings reported in this paper provide no support for the hypothesis that patterns of teacher turnover have a detrimental effect on the quality of public school teaching staffs. In fact, the results support one alternative hypothesis, that there is selective attrition of teachers perceived by their supervisors to be unproductive. All of the selective attrition that does occur takes place during the first years on the job, a pattern consistent with the view that important new information about the quality of the job match is generated during the initial years on the job, and this information causes administrators and/or teachers to take actions that influence the duration of the job match.

FOOTNOTES

1. There have been studies that have related attrition to the training and demographic characteristics of teachers (e.g., Charters, 1970; Greenberg and McCall, 1974; Murnane, 1981; Schlechty and Vance, 1981). However, since these variables explain little of the variation in teaching performance, studies of this type are of limited value in investigating whether turnover systematically affects the average quality of the teaching stock.

2. The scores of the 1,545 second and third grade students on Metropolitan Achievement Test batteries were standardized to eliminate differences in the distributions across grade levels. This was necessary to make the sample size as large as possible.

   See Boardman and Murnane (1979) for reasons why the specification in equation (1) is preferable to a first difference specification.

3. The vector of child and family background characteristics included:
   1. a dummy variable for a male child,
   2. a dummy variable for living in publicly subsidized housing,
   3. the percentage of the population under 18 years of age on the block who lived in a female-headed family,
   4. the number of days the child attended school during the school year.

4. The values of the F statistics were 2.22 and 3.19, for reading and math instruction respectively. These values are sufficiently large to reject the null hypothesis at the .01 significance level.

5. Prior work with these data indicate that different principals use approximately the same standards in evaluating teachers. There is evidence (Murnane, 1975, p. 52) that differences across schools in the

average evaluations of teachers reflect differences in average performance rather than differences in the evaluation standards used by principals.

6. For 72 percent of the 104 teachers in the sample, performance was assessed during the 1970-71 school year; for 14 percent of the sample, performance was assessed during the 1969-70 school year; and for 13 percent of the sample performance was assessed during the 1971-72 school year.

REFERENCES

Boardman, Anthony & Richard Murnane. Using panel data to improve estimates of the determinants of educational achievement. Sociology of Education, April 1979.

Charters, W.W., Jr. Some factors affecting teacher survival in school districts. American Educational Research Journal 7, 1970, 1-27.

Greenberg, David & John McCall. Teacher mobility and allocation. Journal of Human Resources 9, Fall 1974, 480-502.

Hanushek, Eric. Education and Race. Lexington, Mass.: D.C. Heath, 1972.

Jovanovic, Boyan. Job matching and the theory of turnover. Journal of Political Economy 87 (5), October 1979, 972-990.

Medoff, James L. & Katherine G. Abraham. Are those paid more really more productive? The case of experience. Journal of Human Resources 16 (2), Spring 1981, 186-216.

Medoff, James L. & Katherine G. Abraham. Experience, performance and earnings. Quarterly Journal of Economics 95, December 1980, 703-736.

Murnane, Richard J. Teacher mobility revisited. Journal of Human Resources 16 (1), 1981, 3-19.

Murnane, Richard J. The Impact of School Resources on the Learning of Inner City Children. Cambridge, Mass.: Ballinger, 1975.

Schlecty, Phillip C. & Victor S. Vance. Do academically able teachers leave education? The North Carolina case. Phi Delta Kappan 63 (2), October 1981, 106-112.

Weaver, W. Timothy. Educators in supply and demand: Effects on quality. School Review 86, August 1978, 552-593.

Essays on School and Policy Analysis

Essay 1: Teachers Control Students, Students Control Teachers

(This essay is a revision of a briefing to senior staff members of the National Institute of Education, May 22, 1980.)

by

Edward W. Pauly
Institution for Social and Policy Studies
Yale University
1982

This research was supported by the National Institute of Education under grant NIE-G-79-0084. NIE is not responsible for the content of the research.

Historically, the question of how school finance is related to student achievement has led a number of people to search for stable relationships between resources and student achievement; to search for stable relationships between federal school finance policy and student achievement. I think we have learned a great deal from that work.

But one problem with it, a problem that I think comes from the assumptions that underlie it, is that it has a tendency to look for answers in the resources themselves, in the curricular programs themselves, and in the policy actions themselves, rather than in the seemingly unstable relationships among teachers, students and policy makers out in local public school systems.

The question I want to ask is--if we want to improve the way the work of learning is carried on in schools, what do we need to know about how the people in the schools actually behave, actually get their work done? How can we understand the school behavior that produces learning, and how that behavior is shaped and influenced?

Let me say a few words about my methods in this work. Instead of synthesizing and reinterpreting a broad body of empirical studies, I have concentrated on developing new conceptual models to clarify our thinking about how classrooms produce educational achievement. I have been grappling with puzzles and inconsistencies in the literature, and building up a set of ideas with which to approach those puzzles. So it is models and new ways of looking at things that I will be talking about today. And I want to remind you that I will not be speaking about documented findings or empirical results. We feel that later empirical research will make a lot more sense once we have re-thought the assumptions underlying it.

In the model that I'm going to talk about, I begin with the idea that what goes on in a school is up for grabs, and that the main business of teachers and students is to make something of schooling for themselves. In doing so, they struggle with each other. Their struggles with each other determine how life in each classroom in every school will be lived. I don't mean that teachers and students are especially aggressive or uncooperative, but I do mean that they care about what happens to them, and they act strongly to influence what happens to them.

In these classroom conflicts, everything is at stake: the terms of work, the ways gratification will be allocated, people's membership in social groups, and whether help and support will be available. And, it seems to me, these questions are worked out and settled in each individual classroom. And all of these things are contested among a group of people who did not choose to be with each other; they were sent there, to work things out for themselves.

When I say that the business of the school is up for grabs, I mean that people in classrooms decide for themselves what to like and what to dislike in school. Teachers and students try to limit their involvement in what they don't like and they act to get more of what they want.

What is there to like and dislike about schools? What is really at issue among teachers and students?

To begin with: who the other people in the classroom are. Whom must I get along with, simply because they are in my classroom? Are they black, white or brown? Poor or affluent? Like me or different? Do I know them, and are they friendly to me?

Next: how will these strange people treat me? Will they give me a chance to do what I want to do? To excel? To make friends? Or will they hurt me, reject me, humiliate me in the eyes of people I care about? What if I need help--will I get it? From whom?

How will we spend our time together in this room? Will I have any time to myself, any privacy? Will it be boring? How much math will there be? How much homework? Will there be time to talk to my friends? Do I have to pay attention to the lesson all the time?

What happens when things go badly? Are there ways I can get protection, solace, revenge, or just a fair shake when I need help? How painful are the inevitable problems, mistakes and conflicts going to be? Do they go away, and what scars will be left afterwards?

I want to suggest that these questions are not only of critical importance to the students in the classroom; I would say that each of them applies with equal force to the teacher in the classroom.

No one in school is indifferent to these questions. These questions make the difference between seven hours a day becoming happy and full and productive, or draining, empty and painful. For teachers and students alike, these issues take top priority. They are worth fighting about because their answers determine the conditions for life, for growth, for self-expression, even for safety and protection, the things that surround every person in the classroom.

Teachers and students struggle among themselves to make the choices that answer the persistent and unavoidable question of the classroom: how each person will be treated by the others. These struggles determine the ways teachers and students will work together. That's important to our topic today because it is teachers and students who turn school programs into concrete realities, and that takes a lot of doing.

Now I want to speak about how teachers and students get what they want in the classroom. I think most of us would agree that teachers do a lot of things that shape and structure the classroom environment. Simply explaining a new lesson is a time for the teacher to choose how to motivate students, how to gain their attention, how to silence chattering or objections. Breaking up a fight tests the teacher's ability to pick the right moment and the right words to stop the participants, and discourage the onlookers from arguing. When teachers do these things effectively--when they give orders that are obeyed, and explanations that will be believed--they get immediate feedback: life in the classroom goes as they want it to. And if the teacher misjudges, the consequences are immediate, direct and very painful. When students respond to a teacher's demand with an angry silence, the incentives for the teacher to figure out a better way to do things are very strong. And that sort of disappointment makes teachers all the more sensitive to what to do next time.

Now: are students equally capable of acting to make their life in the classroom into something that suits them?

It seems to me that our reluctance to look at how students shape the classroom points to our overwhelming investment in the search for policies intended to make students learn. We think of students as vessels to be filled or as raw materials to be processed. Schooling is often seen as a treatment to be applied to students.

The possibility that students may actually create classroom opportunities for themselves has not received much attention.

Now, it is apparent that students do not suggest course offerings or teaching methods. They have no authority to issue orders, to hire and fire, or to vote. They do not directly put up the tuition required to get into private schools. Younger students may even have difficulty persuading their parents to intervene with a teacher. So long as we focus on these official definitions of what students do and don't do, we will only be able to see students as reacting rather than themselves acting, choosing, bargaining.

I want to suggest that students actively work to get what they want in classrooms. Students can bargain, negotiate and seek compromises in return for their compliance and cooperation. In requests for help and in complaints about an assignment, students implicitly offer to cooperate, and even to contribute their own resources of enthusiasm and commitment, if an agreement that suits them can be worked out.

Students are often able to see exactly when it is hardest for teachers to enforce the rules—when it is most costly in time and costly in emotional energy--and their ability to choose when to hold out for changes in the life of the classroom is a powerful strategic advantage.

Even when they are not aware or mature enough to engage in this sort of give-and-take about their treatment, students can resist and refuse to participate in classroom activities.

Student recalcitrance is effective partly because it can be ambiguous. It is very difficult to know whether somebody isn't paying attention because he can't understand, is ill, or is passively resisting the teacher; it gets even harder when a whole group of students is inattentive. Even the youngest student can choose this sort of ambiguous behavior--behavior that makes equal sense as a child's short attention span, a childish temper, or strategic withdrawal from an unpleasant setting.

In other words, students take advantage of the strategies of the weak. They can disrupt the class, which allocates the teacher's time to crowd control; they can use academic performance to reward a teacher's effort, and even to validate the teacher's sense of adequacy. In effect, I am suggesting that students have the power to contest and control what goes on in the classroom. Teachers control students, students control teachers.

People in classrooms worry about a lot of things, but one thing is on everyone's mind: the question, will I be allowed to do what I want to do? That goes for teachers and students alike. This means two things have to be worked out--the opportunities for gratification that will be available to people in the classroom, and the protections that they will have from others. These two things--chances to do what you want, and assurances of safety in which to do it--are important to all teachers and students; equally important, they are things that all teachers and students can do something about. They can do something about these things because each person in the classroom depends on and needs cooperation from the others. Teachers need the assistance of students, and students need help from the teacher and from each other. And those needs, that relationship of dependence, means that each person can do something about getting what he or she wants from the others.

Students want the assurance that they will have ways to "make it" in the peer culture. Getting along in school, being accepted by classmates, and being able to work for recognition and rewards--all of these things depend on the provision of opportunities by one's teacher and one's classmates. If those opportunities are not available, the student has very little incentive to cooperate with the others; but when opportunities for making it are available, there is a concrete inducement for a student to contribute to and take advantage of what the classroom has to offer.

Equally important to students is protection from humiliation, rejection or being hurt by others. These risks are considerable for the student (child or adolescent) surrounded by twenty-five relative strangers, of all sizes and skills, and creating innumerable demands and pressures themselves. For a student not to be intimidated into withdrawal--and thereby losing the oportunity to make it among peers--requires a guarantee that the threats that come from other people are safely regulated. So students are dependent on the teacher and their classmates for that assurance of safety.

Teachers need to have the opportunity to be helpful, to be supportive, and to get some acknowledgement from students of their efforts to be helpful. The basic gratifications of teaching depend on having a "giving and getting" relationship with students, and that can be destroyed when students withdraw, or are inattentive or rejecting. When students cooperate with the teacher, there are innumerable opportunities for a teacher to have personal contact and recognition. And . . . since teachers require something that only students can give, they are quickly driven to pay attention—not just to curriculum, but--to what students want.

The other factor in teachers' ability to get what they want in classrooms depends on outsiders. By outsiders I mean principals, parents, specialists, officials; people who aren't members of the classroom. School policies, visits from the principal, required curricula and parents' interventions all can undermine the carefully worked out and struggled over arrangements among teachers and students. Now, teachers are responsible to these outsiders, and vulnerable to them. But after the outsiders have departed, it is the teacher who has to resolve the conflict between what the principal wants and the needs of a student who would be better off with, for example, some other instructional approach. Out of this conflict comes the teacher's need for privacy-- the protection of the classroom as a relatively independent place, a place that is safe from the demands of outsiders who don't have to live with the disruption that they cause.

Teachers get this privacy by maintaining the fiction that everything is under control and that all school policies are in full force--something that is probably never the case. But for this to work, students have to go along--by keeping quiet. This creates another strong tie between teachers' and students' treatment of each other. Teachers depend on students' cooperation to be able to have a sense of classroom privacy.

I have been describing teachers' need for privacy as if it were solely a response to some nasty, threatening bunch of outsiders. I think it is true that outsiders' understanding of classrooms will inevitably be partly incorrect; but there are other reasons, purely internal to teachers, that cause them to seek privacy. The problem is that no teacher can be fully adequate to the needs of twenty-five or more children. Consequently, teachers ask themselves, what if this inadequacy were to be discovered? Seymour Sarason has called teachers' perception of their own inadequacy a central dilemma of schooling. Because public school teaching seeks to provide every child with appropriate instruction, protection for the teacher--that is, privacy--privacy in which to cope with the occasional or frequent disappointments is a necessity. And in order to have that privacy, teachers need students' help.

I want to suggest that teachers' and students' needs to gain these opportunities and protections from each other allow them to be controlled by each other. And the quality of classroom life depends directly on how they resolve their struggles. The struggles over protection, and over opportunities for gratification are thus a regular, predictable and central part of life in classrooms. And what are the outcomes of the struggle?

The specifics will surely be highly variable. In fact, the main characteristic of students' wishes about how classroom life should be is how changeable and uncertain they are. Their desires for a balance of work and play, their choice of one course over another or which activity to engage in depend on their moods, their friends, the teacher's enthusiasm; and these wishes are relatively easily altered. And teachers, too, are often uncertain about what will give them the satisfactions of having been helpful, or stimulating, to their students. The result is enormous variation in the way classroom tasks are carried out, variation that tends to disguise from us the unceasing and, I think, the consistent struggle among students and teachers to sort out who gets what in the classroom.

I want to think through the implications of the particular kind of struggles that I have just described for classrooms. It seems to me that through these struggles, the teachers and students in a classroom work out how they will treat each other, and consequently how the work of learning will get done in the classroom. In other words, one could say that through their classroom struggles, teachers and students govern what goes on in the classroom; they direct it, organize it, set limits on it, make authoritative decisions about classroom life and work.

And classroom governance has certain predictable features. The result of the struggles that involve teachers' desire for giving and getting relationships and students' desire for opportunities to make it in the peer culture is a collective classroom agreement, an agreement that personal responses will generally be made available to each person in the classroom. In other words, everyone, teacher and students, can get attention and personal exchange in some fashion in the classroom. The particulars of this responsiveness depend on the specifics of each classroom's struggle; in some classrooms the agreed upon personal responses will be ones we like, in other classrooms ones we don't like. But every classroom's agreements contain some sort of settlement in this general category.

The result of the struggles that involve teachers' desire for classroom privacy and students' desire for protection from hurt is a collective classroom agreement to limit each person's treatment of the others. Again, the particular limits depend on the specifics of each classroom's struggle. The results will be different in different classrooms, but every classroom's agreements will have something to say about limits.

These two outcomes of classroom governance establish a framework for the life of the classroom. I would go so far as to label them the constitution of the classroom. That's not a conventional use of the word constitution; I think it suggests the importance of a practical and highly tailored set of "rules of the game," rules that hold people together and simultaneously allow room for the autonomy that individuals need to learn and to teach. What I am saying is that people in classrooms govern themselves; they make their own constitutions. When I say that the constitution creates a framework for classroom life, I mean that it can allocate status, it can reward and value particular kinds of work, and it can regulate the personal and emotional contributions to classroom life.

The agreements, the constitutions, that people in classrooms make allow them to draw on each other--to get help, to feel safe, to find encouragement and support. Consequently, the whole becomes much more than the sum of its parts. The human resources of the classroom, freely drawn upon, can make public school classrooms rich in resources, not poor, when there is a supportive constitutional arrangement.

If my description is right, classroom governance can be said to have two important results: first, it greatly influences the substance of schooling in the classroom: the organization of its tasks, its schedules, its rewards, the quality of its relationships, its ability to stimulate its teachers and students to participate, and its sustained priorities. Second, the phenomenon of classroom governance means that outside policies and mandates will not be passively accepted--instead, they will themselves become the occasion for conflict over how people will get along with each other in each particular classroom.

It seems to me that teachers and students carry out some extraordinarily creative and powerful processes in governing their classrooms, and to the extent that we have missed the significance of these processes we have blinded ourselves to the potential of some enormously productive sources of motivation and learning.

The last thing I want to say about this line of thought is that I think it helps us make sense out of some very puzzling observations that people in education have been troubled by.

One is the fact that apparently successful school innovations have not been successfully replicated; it may be the case that these classroom interventions are easily dominated by the internal governance of individual classrooms.

Another, one that Dick has discussed, is the finding that students in different classrooms make different achievement gains, even when we control for the extremely powerful home and background factors that students bring with them, and these classroom differences are not primarily due to measurable resources. What accounts for this differing productivity of classrooms, even classrooms in the same school? I believe that much of the answer lies in the way that the classroom constitution shapes the way that teachers and students choose and carry out learning tasks.

A final example: educators, and surely every parent, have often observed the phenomenon of a child learning all by himself--suddenly discovering some previously ungrasped concept, with no teacher or school program in sight. Perhaps one little-noticed but powerful way that schools do contribute to learning is in the social rules worked out in classrooms that remove blockages that preventing students from learning by themselves, blockages that limit energies and motivation, or define success and failure in troublesome ways. Such blockages as these may be undone by classroom rules that protect students from threats and hurts, and by rules that reward and value work and learning.

These speculations aren't new, I would say, but they are strong suggestions that our efforts to relate school finances to student achievement need to pay close attention to the classroom processes of production and control.

First, we should not expect top-down control efforts, or subsidies, or mandatory fiscal regulations on the behavior of people in local schools to dramatically improve student achievement. These initiatives, including fiscal controls, will run up against the highly idiosyncratic and strongly defended constitutions of two million classrooms across the country, but I would hope we can begin to regard them as a source of strength rather than a barrier.

Second, and more important, I would suggest that the school tasks that we should increasingly be concerned about are not ones that are supposed to make teachers and students learn, or even stimulate them to learn, but the ones they choose for themselves. If they choose learning tasks that fail, we need to ask why, not just impose our solutions on their classrooms. If they choose tasks that succeed, we need to support them even if we can't figure out how they do it.

What difference does this work make?

It should make us look hard for ways to take advantage of the often unnoticed power of the key human resources of schooling, teachers and students. Their behavioral responses and their classroom governance are sources of motivation and learning that can be enormously productive. We believe that the roots of the next generation of school finance policy lie in an understanding of these processes.

Finally, we would say that in attempting to understand the relationship of school finance to educational achievement, we should pay close attention to the various and subtle behaviors of the crucial human actors who actually link the dollars to the results. This is what we have begun to do.

Bibliographical Note for Essay 1

Many of the ideas in this essay have drawn heavily on Sarason, The Culture of the School and the Problem of Change (Allyn and Bacon, 1971) and Sarason, Levine, Goldenberg, Cherlin and Bennett, Psychology in Community Settings (Wiley, 1966). For an earlier example of a social scientist's careful analysis of the workings of conflict among purposive actors, see Schelling, The Strategy of Conflict (Harvard University Press, 1960). An analysis of conflicts that emphasizes the apparently weaker party is Janeway, Powers of the Weak (Knopf, 1980). The cited report on the non-replication of so-called "lighthouse" school innovations is The Ford Foundation, A Foundation Goes to School (1972). Descriptions of the variation in student achievement in different classrooms can be found in Murnane, "Interpreting the Evidence on School Effectiveness" (Teachers College Record, Fall 1981), which cites and analyzes all four relevant studies.

Essays on School and Policy Analysis

Essay 2: On the Political Nature of Classroom Teaching and Learning

by

Edward W. Pauly
Institution for Social and Policy Studies
Yale University

1982

This research was supported by the National Institute of Education under grant NIE-G-79-0084. NIE is not responsible for the content of the research.

People who live and work with each other quickly discover that the way they can live their lives depends very directly on the way others treat them. So they often make agreements, typically unspoken ones, about how to treat each other. Because these agreements are extremely important, they are struggled over. Struggles and conditional offers of cooperation are at the heart of the approach I will use to explore how students learn in public school classrooms; for conciseness, I call it a political approach to studying how schools work. The approach is political because it examines the processes by which people in classrooms govern their relationships with each other, and the consequences that these processes of pressure, influence and adjustment have for teaching and learning.

I will suggest that political processes play a significant role in learning and teaching. How can this be? The extensive studies of learning behavior carried out by experimental psychologists emphasize the dependence of learning on such processes as repetition, reinforcement, and step-by-step mastery of new skills. Our views of learning have thus been shaped by the image of the laboratory rat running a maze, learning its twists and turns in pursuit of food. That view has supported the development of teaching materials and techniques for improving the transmission of knowledge from the experiment-designer (the teacher) to the experimental subject (the student). But consider another possible metaphor for studying learning: what if the first experimental psychologist had put two rats into the maze, instead of one? Perhaps they would have paid a good deal of attention to each other, and the nature of their interaction might have been a crucial determinant of whether and how they pursued any given goal set by the researcher.

Classroom learning is carried out by students and teachers who find themselves in the maze, and surrounded by other people.

Consider for a moment some of the ways that learning goes on in classrooms.

-Students compete with each other to see who can finish the assignment first, or who can get the teacher's attention.

-A student having trouble with a math problem is helped by the teacher, sometimes effectively and sometimes not; or is sometimes reprimanded.

-The teacher may become caught up in the class's unexpected excitement during a discussion of what the first European explorers in America found, and be stimulated to extend and revise the planned lessons.

-The teacher, uncertain about how to motivate students in a writing project, searches for ideas by recalling conversations overheard among students, about people they admire and events that have excited them.

These social behaviors are exchanges among people in classrooms that influence how they go about teaching and learning. Such exchanges, unpredicted and largely beyond the control of any single participant, go on constantly in classrooms, simultaneously with whatever learning or failures to learn make up the day's educational accomplishments. People in classrooms are social in their behavior, though in widely differing forms and degrees. What is constant is the tangled relation of learning with the ordinary processes by which people get along with each other.

But the useful exchanges that help teachers and students achieve together do not always go on. When student or teacher finds it too costly, or risky, or confusing to make an agreement with someone else, their exchanges can become aimed at guarding or separating themselves from others.

Agreements that support teaching and learning do not happen by accident. They must be hammered out, and only when it is to the advantage of a group of people to hammer them out will these agreements be reached. I want to suggest now, and in the remainder of this essay to explain, that the success or failure of the public school classroom as a setting for learning is drastically influenced by the classroom's pattern of personal exchanges and interactions. And this pattern is itself a product of political struggles among the teachers and students in each individual classroom.

Let us look more closely at two examples of the interactions that determine how life will be lived in the classroom: helping and competing. Both helping and competing occur among teachers and students when people choose them as ways to deal with other people. They make their choices of how to treat others as part of their individual efforts to live, each one, as he or she wishes; helping and competing with others are chosen when they make sense for the individual. These choices will make sense, and will be chosen, when they are seen as likely to elicit desired treatment from others.

A student who is offered help may want to accept the help because it provides a way to solve a confusing problem, because of a wish to receive attention, or for many other reasons. But at the same time that student may fear being subordinated by the helper, dominated, or even coerced into unwanted acts. It should not surprise us if that student refuses help. And we should also anticipate the student's efforts to constrain and control the help that is offered. If he or she can place limits on the kind of help, or on the terms of acceptance--if the student can partly control the rules of the helping transaction--then it makes sense to join in.

The helping person, however much an altruist, may also appropriately reckon with the possibility of being exploited--for a quick correct answer, or for a continuing supply of free labor. Thus the helper, too, chooses whether to withdraw from helping or to control, in part, the rules of the encounter. Struggles over the control of the helping relationship provide a striking example of the dependence of teaching and learning on "political" relationships in the classroom.

Even the ways that teachers teach are subject to these--call them what you will--controls, struggles, or simply settlements of emotionally charged exchanges. Teachers depend heavily upon their knowledge and their informed guesses about what students' responses to their actions will be. The knowledge that teachers need is intimate, detailed and personal knowledge about particular students: their preferences, their thoughts, their relationships with each other and their commitment to activities and norms extending well beyond the classroom door. Every effort to explain what a literary symbol is, every attempt to gain students' attention when a discussion turns off point, and every choice of encouraging words to a slow and unhappy student relies on this knowledge, and is likely to dissolve into confusion, incomprehension and disorder without it. Students' messages to a teacher have the effect of constraining and molding the teacher's knowledge about them, thus partly controlling the way the teacher teaches. The availability of the knowledge required for teachers to teach is controlled by students.

Thus far we have discussed how teaching and learning are shaped by struggles over the terms of helping relationships (and over the knowledge required to sustain such relationships). Competition, along with other expressions of aggressive behavior, may also have the effect of stimulating learning in the classroom. School achievement can be a marker in a game that operates on several levels at once. Competing is a way of gaining other people's attention, whether by showing off or by fitting in. When students want opportunities to strengthen bonds with their peers, competing can be a way to transform an isolating experience in the classroom into a chance to make a successful connection with one's peer culture. It follows that classroom competitions, like helping relationships, allow people to pursue several goals at once; for example, a student's fair play and display of good form may be at least as important as winning--whether what is at stake is a top grade, the teacher's attention and recognition, or being first to complete a task.

For teachers, there is also a great deal to gain in competitive life. Challenging students to perform beyond their previous achievements, and even goading an individual student to engage in a difficult task, may lead both students and teacher to learn more than they otherwise would. When the norms established by a teacher and students for governing classroom life clash with values from outside (neighborhood, street peers, or even other classrooms), the teacher competes with these other influences; a teacher's successful assertion of the primacy of the classroom's ways of solving problems can contribute to maintaining a setting that encourages achievement, the exchange of help, and other useful interactions.

These forms of classroom competitiveness are powerful molders of behavior, and in particular, of teaching and learning. The strategy of competition depends on paying attention to one's adversary; as a result, that person's actions constrain and partly control one's actions. Classroom learning can therefore be said to be partly controlled by the competitions that arise among teachers and students.

The controls are all the more effective because the very terms of competition are implicitly negotiated among teachers and students. In order for a student to be willing to join in classroom competition, he or she must be confident that the possibility of being humiliated or harmed is controlled, whether by rules of behavior or guarantees of protection. The most often reported, and surely intensely felt, fear of teachers--that they will not be able to gain their students' attention--serves as a strong pressure for the teacher to make sure of ground rules that meet their needs before entering into competitive exchanges. Because teachers and students both need to impose limits on the competitiveness from which they benefit, they act to limit and thus partly control the terms of competition. In doing so, they determine the pattern of learning that can be stimulated by competition in their classrooms.

These two parts of classroom life, helping and competing, are not the only powerful exchanges of behavior that govern and support teaching and learning. Play, discipline, small group work, complaints and criticism and praise, all adapt the tasks of teaching and learning to the tasks of living together with people in the classroom. And in every form of shared classroom life, opportunities for give and take as well as the controls placed by one person on another direct the route taken by teaching and learning.

Can we say, then, that school learning is a product of the daily struggles of teachers and students to get along with each other? Yes and no. Learning is in many ways an individual activity, and the choices teachers and students make in purely solitary ways have a great impact on learning. The motives established in each person by home and neighborhood are social influences on learning that are internalized by the individual and thereby shape learning. But for school learning--the skills and ideas that people discover, learn, practice and apply in the active and crowded setting of the public school classroom--teachers and students depend directly on their interactions with the people around them. Their exchanges with each other dominate the teaching and learning that go on in the schools.

I would go further than this claim. It seems to me that it is possible to examine the patterns of interaction that are worked out by students and teachers in the classroom, and to trace their relationship to students' achievement.

Partly through habit, partly through negotiation, teachers and students work out their joint responses to the uncertainties of classroom life. These uncertainties are very great; students need assurances that they will have opportunities to succeed in their peer culture, and teachers need acknowledgments of their helpfulness. Since each person in the classroom depends on the others for these assurances, the pressure for everyone to join in a shared set of rules is very great. I suggest that the uncertainties and interdependence of teachers' and students' daily lives are an important source of give and take. (Elsewhere I have written about this process of political conflict and compromise, and its results.) This give and take takes different forms in each classroom. But it seems to me that two broad kinds of agreements get worked out in most classrooms: limits are placed on the ways people may treat each other, and the terms and availability of access of each person to the others is made known and established. Because these agreements, whatever their specific content, reflect the basic expectations of people in the classroom of how they will get along with each other, we can call them the classroom constitution. I became interested in the creation of classroom constitutions when I noticed how vulnerable were the wishes and activities of each teacher and student to the others in the classroom; although my description of this phenomenon here relies far too much on assertion to be fully persuasive, I think it shows how an individual's learning often depends on the interactions that take place with other people in the classroom.

The classroom constitution shapes learning activities in two ways. (1) By governing each person's access to others--for support, ranking, attention, imitation, and for the availability of other sorts of personal contact--the classroom constitution allows people to draw on one another. Without an informed and firm expectation of the kinds of responses from others that can be obtained, it is simply too risky to engage in helping activities or competition. When the risks of help and competition rise, the participation of students and teachers in actions that support and stimulate learning cannot be sustained. (2) By placing limits on the ways each person may be treated by the others, the classroom constitution reduces the dangers and the uncertainty of engaging with others in the interactions that foster learning, including helping and competitive activities.

Teachers' involvement in interactions that advance student achievement is as sensitive to the classroom constitution as is that of their students. In order for teachers to sustain the difficult, largely trial-and-error search for instructional approaches that are effective for the group of students in a particular classroom, they must be able to anticipate the range of responses and acknowledgments that will be available from students. The constitution governs the ways that the teacher can draw on the other people in the classroom, thus providing information necessary for the teacher's work. When persistent problems of failure by certain students arise, the teacher's effort to gather the highly idiosyncratic information on learning problems and on "what works" for a particular student depends on the establishment of limits to assure both teacher and student that intrusions will not get out of hand.

What I have just suggested is that both teachers and students depend on the assurances provided by the classroom constitution in the interactions with each other that make school achievement work. But the particular substance of classroom constitutions varies widely, even among classrooms in the same school; constitutions will differ in fairness, in flexibility, and in the acceptability of particular forms of behavior in each classroom. (The differences among classroom constitutions occur in both crucial articles of the constitution, the avilability of personal access to people in the classroom by others, and the limits on the ways people treat each other in the classroom.) It would seem to be the case, then, that the nature and the amount of teaching and learning that go on in a classroom are determined by the social interactions made possible and governed by that classroom's constitution. The constitution creates the range of possibilities from which people choose the interactions that make teaching and learning work.

One result of the influence of the classroom constitution is that distinctive patterns of school achievement will arise in each classroom, patterns of some stability that sustain some tasks and effectively prevent others from occurring. Thus, we would expect to find classrooms in which patterns of sustained failure arise, because failure in student achievement may be more secure for the people in that classroom than the risks that they have experienced in occasions of competition, or help offered to others, or other interactions that support learning.

Other patterns of school achievement can be easily imagined. Ritualized compliance with school authority is such a pattern, perhaps resulting from classroom struggles in which people find it more difficult to respond to each other than to stick to highly predictable forms of obedient behavior as a way of limiting the hurt that might be suffered from more spontaneous exchanges. The borderline-violent, disruptive outbursts that become a pattern when they are accepted as inevitable and predictable make up another pattern that can be produced by people in classrooms, and a pattern that will drastically shape the kinds of learning and teaching that go on in that classroom.

These brief sketches are more of a caricature of the ways teaching and learning are shaped by interactions in public school classrooms than anything else; but it seems to me that the idea that underlies them, the idea that classrooms create their own patterns of work, search for methods, and even achievement, is one to which we need to attend.

If the people in each classroom together create a characteristic pattern of achievement, and do so through a gradual, emotionally charged, and political sequence of interactions, struggles and compromises that no one person can control or even fully anticipate, then our efforts to design school policies and programs and rules that depend on a teacher's ostensible control of the classroom are in trouble. If the patterns of interaction that people work out for themselves have significant consequences for the kinds of materials, activities, and even learning goals that they will be able to make use of, then it may be that any relatively standardized school programs and policies detract far more than they contribute to schooling.

These speculations need to be addressed in much more detail before they can shed light on the predicament of American public education. I begin that process in the chapters that follow.

The portrait of public schools that has been drawn in these pages is of an intense, high stakes struggle, one that presents its members with piercing and continuing problems and uncertainties, as well as with remarkable opportunities. The people, students and teachers, who find themselves in this setting struggle to come to terms with it and with each other. It is this struggle, and the "solutions" to it that teachers and students hammer out together, that make their lives livable (even if sometimes in a drained, beaten-down way). And it is this struggle that gives schools their characteristic shapes and energies--and outcomes. If we appreciate students' and teachers' struggles to control each other and to get along with each other, we may be able to support their efforts in ways that lead to more effective and constructive uses of schools.

For there is an extremely positive lesson in this way of looking at schooling: the people in classrooms are enormously resourceful, and they act resourcefully because they have to; together they make up an extraordinarily powerful resource for education. Helping and competition and the making of classroom constitutions are no accidents--they are the hard-won products of difficult and often painful labors. These products, these patterns of personal treatment that stimulate learning in the crowded, rich, dangerous social and political world of the classroom, can be enormously productive. The remainder of this book examines the problem: how can the energies of people in schools be turned to the tasks of greater school achievement?

Bibliographical Note for Essay 2

For a (somewhat dated) review of mainstream studies of teaching and learning, see Gage, Second Handbook of Research on Teaching (Rand McNally, 1973), especially these articles: Beller, "Research on Organized Programs of Early Education;" Green et al., "Research and the Urban School: Implications for Educational Improvement;" and Della-Piana and Endo, "Reading Research." A more critical overview of the literature, and one that has influenced this essay, is in Averch et al., How Effective Is Schooling? (Educational Technology Publications, 1974).

The ideas in this essay on the role of interaction in school life have been greatly influenced by Lindblom, Politics and Markets (Basic Books, 1977), especially chapter 19, "Two Models." A useful analysis of competition and related forms of social choice and allocation can be found in Schelling, Micromotives and Macrobehavior (Norton, 1978). The ideas in this essay on struggles within classrooms are, generally, to be distinguished from the analysis of Bowles and Gintis, Schooling in Capitalist America (Basic Books, 1976).

Also, see the wide-ranging ideas of Waller, The Sociology of Teaching (Wiley, 1932). I am indebted for the two-rats-in-the-maze analogy to Sarason, in The Culture of the School and the Problem of Change (Allyn and Bacon, 1971).

Essays on School and Policy Analysis

Essay 3: How People in Schools Coordinate Themselves Without Intending To (And Other Features of School Life)

by

Edward W. Pauly
Institution for Social and Policy Studies
Yale University

1982

This research was supported by the National Institute of Education under grant NIE-G-79-0084. NIE is not responsible for the content of the research.

Teachers and students, like the rest of us, respond to the opportunities and the problems that they encounter; they choose their actions so as to make the most of their school experiences. This simple beginning has allowed us to consider how teaching and learning are influenced by the struggles of teachers and students as they work out among themselves just how they will treat each other.

In this chapter, I will suggest that what goes on in classrooms is itself strongly influenced by the actions of everyone in the school. My argument is this: since teachers and students care about what will happen to them in school now and in the future, they attempt to learn how the behavior of the people around them limits and shapes their own prospects; bit by bit, their behavior tends to adjust to that of their schoolmates and colleagues. Teachers and students in different classrooms thus achieve a kind of rough, unofficial coordination with their colleagues and schoolmates, and they do so without guidance or sanction from school officials. I will suggest that this takes place for the same reasons that motivate the intimate struggles over the classroom constitution that I described in chapter one: the self-interest of students and teachers. But self-interest works differently in schools than in intimate settings such as classrooms; and in this chapter I will explain why certain distinctive relationships are built up by the teachers and students in a school, and how those relationships work.

In classrooms, each person is known to the others personally and intimately; by comparison, schools are far more impersonal places, where people are often known to each other according to school labels: Jill's older brother in the fifth grade; the strict third grade teacher; my last year's English teacher; that troublesome boy in the seventh grade. In classrooms, life is lived immediately and directly: students obey the teacher's order, or they disobey it; the teacher answers a student's question or ignores it; the math lesson ends in confusion or with a sense of accomplishment; teacher and students treat each other in ways that express boredom, or caring, anger, suspicion, playful camaraderie, or intense competition. But life among the people in a school, beyond the intimacy and directness of classroom relationships, unfolds slowly and tends to be lived indirectly, by hearsay and fragments. The relationships of teachers with their colleagues, and students with their peers, emerge gradually as a large number of observations and exchanges are added up, turned over in people's minds, and interpreted. A new teacher slowly learns that information about problematic students is only available from other teachers when the new teacher endorses the past handling of such students. Students may gradually discover what combination of socializing and schoolwork will satisfy both the students' peers and his or her present and future teachers.

With neither the intimacy nor the directness of classroom life to help them in coping with the school world beyond their classroom door, how can teachers and students decide how to respond to the life of the school, what they want from their school, and how they might act to get it? How do people work out arrangements that determine how they will live together and get along, when they cannot engage in face-to-face interaction and bargaining with one another?

Face-to-face interaction and bargaining do not exhaust the means of people who need to come to terms with each other. Teachers and students, like the rest of us, often decide what to do in complicated situations involving other people simply by learning about those people, and adjusting their own behavior accordingly. People in schools have many opportunities for doing just that. They collect information constantly from each other: information about what things are like in other classrooms; about particular teachers (mean, strict, kind, "good" or "bad") and students (brassy, quiet, crazy, or "the teacher's pet"); and about the consequences in other classrooms of particular behaviors (which actions get punished, which actions stimulate "acting up"). Teachers and students can use the information they get from schoolmates and colleagues to decide how others are likely to treat them, what they likely responses will be to their own acts, and consequently how they can best get along with their fellows in schools.

In other words, teachers and students adjust to what the people around them are up to. And the essential difference between interactions in the school and interactions in the classroom comes down to that: in schools, people adjust to information about colleagues and schoolmates, rather than directly struggling or collaborating with those people. Instead of trying to govern themselves and others (as in classroom life), people in schools adjust to the life of the school around them. The result of everyone's participation in this process of adjusting to each other's behavior is that the people in a school become roughly coordinated with each other. This happens even though people in a school (unlike the people who find themselves together in a classroom) typically do not attempt to influence each other's behavior.

But information about the people and events in one's school is not by itself sufficient to determine what is in one's self-interest. In order to make sense out of information about school life, teachers and students need to condense it, interpret it, make it useful for themselves. This they do in two ways: by making comparisons between themselves and others; and by anticipating how in the future they will be treated by the teachers and students in their school. These two self-interested acts of teachers and students, comparing and anticipating, serve to invoke and transmit the powerful influence of schoolmates and colleagues into every classroom in a school.

Comparison and anticipation, adjustments to school life

As teachers discover, through casual conversations, how much of a reading or math textbook their colleagues have covered, they immediately compare what they hear with their own accomplishments--and wonder whether they have done enough. As they walk past other classrooms, they notice whether the noise level exceeds that of their own classroom. And in trying to discipline a child who repeatedly upsets the classroom, teachers compare their own adequacy to what they imagine to be that of the other teachers in the school. Teachers relate their performance to that of their peers; by comparing themselves to the other teachers in their school, they learn what is possible and how well they are doing at achieving what is possible.

Students also compare themselves, but with schoolmates in other classrooms. By finding out about life in other classrooms in their school, they discover how far they have gotten in a common textbook; what punishments they suffer that others do not; whether their assignments are relatively onerous compared to the assignments their schoolmates get; how much fun their classroom is compared to others'; how mature, or how babyish, their classroom behavior is compared to behavior in other classrooms; how their teacher compares--in "niceness," strictness, helpfulness, accessibility. These discoveries leave students with a sense of what the attainable aspirations are in student life.

When teachers and students compare their classroom experiences with those of colleagues and classmates, they tend in many small ways to make adjustments in their own behavior. As teachers and students compare their situation with that of other classrooms, they may feel deprived or overadvantaged--and their demands on each other may sharpen or diminish as a consequence. For example, a teacher's worry that he or she is falling behind what other classrooms have covered, or students' belief that they are missing opportunities that are available to schoolmates, can cause teachers and students to seek changes in their own classroom. Comparisons also broaden, and even determine, the menu of alternatives that the people in each classroom are aware of. When teachers imitate their colleagues' ideas for dealing with parents' dissatisfactions, they have been influenced by a comparison of their own coping strategies with those of their colleagues.

But teachers' and students' central response to their information about schoolmates and colleagues is to create performance standards for themselves. Teachers evaluate their pacing according to their colleagues' comparative location in a textbook. They establish expectations and aspirations for themselves--and for their students!--by watching neighboring classrooms. And students evaluate themselves in terms of the standards implied by schoolmates' anecdotes about tests, homework, activities, and the acceptability and consequences of various forms of behavior. High school students are acutely aware of tracking--which classes are "advanced," "slow," or plain "bad." Even the youngest students learn performance standards that label certain inappropriate behavior as "babyish," but what this standard is will vary from school to school.

Comparisons tend to adjust the people in each classroom to all of the other people in the school. Sometimes classrooms in a school become fairly similar (in the pacing of lessons, for example); sometimes different classrooms adopt different standards, and defend their standards by rejecting what goes on next door. But in either case, teachers and students respond to comparisons with other classrooms. Their response is, of course, a self-interested one and one that reflects how the people in each classroom choose to get along with their schoolmates and colleagues.

Teachers and students also use information from people in other classrooms to anticipate how they will be treated in school in the future. When a fifth grade teacher hears of problems in fourth grade classrooms, he or she thinks in terms of what those problems might become next year—when the fourth graders become that teacher's fifth graders. And when students hear that their schoolmates in the next grade are expected to be able to write short research papers with footnotes, their anxiety is the result of anticipating what such an assignment might mean for them. In school, visions of one's likely future are quite accessible; so teachers and students are able to see how they are likely to fare when the experience of their colleagues and schoolmates has become their own.

People in schools use their anticipations to calculate how they can affect their uncertain future by what they do now; they adjust their behavior to avoid problems they would otherwise anticipate. When teachers race to finish the textbook so that their successor won't criticize them, or when students choose their fall wardrobe so that they will fit in with new classmates, they are taking action now to stake out their preferred version of school life in the future. Anticipating the future consequences of one's actions requires guesswork, but the guesses are informed by a great deal of information from schoolmates and colleagues: by observation, by conversations, by asking hypothetical questions, and by calculating actions to avoid the undesirable consequences that one anticipates. Many of these calculations are aimed at self-protection: when teachers act to protect the privacy of their classroom, to maintain supportive relationships with colleagues, or to deter problems from a student whose anticipated behavior is disruptive, they must first use their information about school life to anticipate problems, and then calculate what difference various actions might make for the feared problems. Students protect themselves from anticipated problems, too: they avoid doing things that might disrupt friendships with peers, or might lower the achievement levels they expect, or might antagonize powerful strangers (such as their likely future teachers). For teachers and students alike, the information that they need in order to anticipate the consequences of what they might do cannot be found inside their own classrooms, but must be obtained from the schoolmates and colleagues they encounter outside the classroom door.

All of this activity has the effect of getting the teachers and students in a school to adjust to each other by paying attention to each other's behavior (and thinking about some of the reasons for it), and responding accordingly. What I mean when I say that they will respond accordingly is that it is strongly in the self-interest of all teachers and students both to try to anticipate how they will be treated by others in their school, and then to choose particular behaviors in light of what they have learned from the other people in the school. Sometimes students and teachers adjust to what they anticipate by conforming to what others demand of them, and sometimes this response is to do the opposite--but always they adjust, and the framework for their adjustments is created by the climate of anticipation that is transmitted, discussed and used by all of the people in a school. Because the people in a school depend on each other for the information that they need in order to anticipate how they will be treated in the future, and because their information about school life is commonly-shared, their anticipations draw the people of a school into a continuing, self-reinforcing process of adjustment to each other.

Students and teachers adjust to their anticipated school experiences in some highly predictable ways. They get prepared for their next classroom. (Teachers collect lore about students they will teach in the future; students focus on the skills, social and academic, that will be needed in their next classroom.) They learn what choices will be available for them, and what difference their choices will make; and they either reconcile themselves to this range of alternatives or, more rarely, think of changing it. (Students pay great attention to the menu of courses and teachers available for them to choose from; teachers worry about limitations imposed by available textbooks, or the boundaries of the official curriculum.) And each response of a teacher or student to the information about what is likely to happen to them in school adjusts that person's behavior to take account of the shared school experience.

The adjustments that teachers and students make in their lives as a consequence of their comparisons and anticipations often have powerful consequences for the teaching and learning that go on in a school. Teachers' responses to their students' problems are shaped by the lore from other teachers about what to expect from that student and from the school's problem students in general. When teachers exchange lore about students, they are also implicitly comparing their experiences and their feelings of adequacy; they are also anticipating how their colleagues will judge them through the behavior of their past students, now in other teachers' classrooms. When teachers decide to conform to each others' behavior, when they rate and rank each other, and when they decide whether and when to help their colleagues-- all of these behavior arise through the piecemeal accretion of teachers' comparisons with each other, and their anticipations of future treatment. When they calculate how to improve future relations with colleagues, it is their comparisons and anticipations that guide their actions. Their teaching decisions and their responses to students are partly guided by their adjustments to the other teachers and students in their school.

Students' classroom lives are shaped by the same school relationships. Self-comparisons and the anticipation of future school experiences provide a powerful incentive for students to keep up and get prepared (or in some cases, to give up and abandon preparations that appear to be worthless). Students notice the effect of school achievement on how their older schoolmates are regarded and treated; their sense of the value of classroom efforts is greatly determined by the buildup of school comparisons and anticipations. So students, too, adjust their learning behavior in light of the school world of which they are a part.

As teachers and students incorporate their self-comparisons and anticipations into their classroom lives, the different classrooms in a school gradually come to adjust to each other. Disparate classrooms come to focus on certain expectations, certain limits and pressures to conform, and a menu of possible alternatives for teaching, learning, and getting along together. This shared focus may lead to conformity across classrooms or conflict and dispute among them, but it is a predictable focus, sustained in everyone's attention by the self-interested anticipating and comparing that teachers and students in a school engage in.

Repetition and sequence in the organization of school life

What accounts for the predictability and the sustained quality of teachers' and students' adjustments to the life of their school? It seems to me that teachers' and students' ability to use information--comparisons and anticipations--from colleagues and schoolmates as the basis for adjusting to their present and future school lives depends on the very special repetitive and sequential nature of institutionalized schooling. By repetition, I mean that schools perform essentially the same sequence of tasks each year. The scenes of schooling are recreated each September; the same grades and courses, with a cast comprising many of the same teachers and students, are assembled for a revival of a familiar (though differently produced) play. The sequential nature of schooling means that each person in the school moves through an established, well-known repertory of roles. Students move from grade to grade and from class to class; they know where they are in the sequence and where they will be in the future, and they can identify all of the other people in the school according to their place in the sequence (and therefore when the student will encounter them). For teachers, the sequence is one of class succeeding class; each year a teacher can observe the students he or she will have next, and the students from last year who have gone on (carrying the teacher's stamp with them!) to their next classroom, their next teacher. Teachers and students identify themselves by their roles in the repeating sequence of schooling: I am the tenth-grade math teacher, or the fourth-grade teacher; I am a freshman in high school, and next year I will be a sophomore here; I know whom I will be with, and when, and doing what, tomorrow and a year from tomorrow.

The repetitions and sequences of school life make it possible for teachers and students to draw meaningful comparisons between their own experiences and those of other classrooms. The fight yesterday was in the same grade level as their own (or in a class two years ahead of theirs); it involved the math teacher who teaches all the geometry classes. The meaning and impact of a comparison depends on how transparent it is, how accessible to people in many classrooms; the repetitive, sequential nature of schooling means that most people in a school know a great deal about what goes on in every grade and course, and regard it as significant for their own lives. As a result, comparisons become highly accessible warnings to the teachers and students of a school--and they tend to be taken as harbingers of the future. If one teacher is criticized this year for low test scores, the other teachers in that grade are likely to worry about their vulnerability to the same problem.

The repetitions and sequences of schooling also underlie the great significance of anticipation in the school lives of students and teachers. This year's eighth graders are next year's ninth graders--so everything that happens in the ninth grade is powerfully salient to them. Likewise, the fifth-grade teacher knows that this year's fourth graders are just around the corner, and rapidly approaching. The step-by-step progress of school grades and courses brings a clarity to teachers' and students' anticipation that is lacking in many other institutions. Because they know in substantial detail where they will be next year--and with whom, and doing what--people in schools are powerfully guided to attend to the experiences of schoolmates and colleagues whose present contains their own future.

All of this suggests that teachers and students see themselves as having a career in school, and they see their self-interest as stemming from a long chain of possible experiences, whose concreteness and availability are demonstrated by all of the people around them in their school. Step-by-step progressions, highly visible repeated events, the anticipation of future experiences, and careful comparisons with the experiences of one's peers--these are all careerist ways of thinking and behaving. This is not to suggest that the lives of teachers and students should be perceived as being purely careerist; to the contrary, the distinctively different way of life that goes on within each classroom keeps teachers and students from behaving much like junior executives. But in another sense, each school does evoke career-like processes in schools--and that is the sense of organized, interrelated behavior and attentiveness that permeates the life of a school. Simply because of the situation they find themselves in, the teachers and students in a school compare themselves to each other, anticipate how they will fare, and without anyone forcing or even suggesting it, they adjust their actions to take into account the acts of their colleagues and schoolmates.

The adjustment process and its result: coordination

Within each school, teachers and students change their behavior as they adjust to the influence of their colleagues and schoolmates. These changes in behavior, driven by the constant self-comparisons and anticipations that characterize school life, affect what goes on in every classroom: the performance standards, choices of activities, allocations of time and effort, and classroom members' expectations and preparations for future classrooms. As teachers and students in each classroom react to the influence of the school, their expectations and their behavior come to cohere with those of people in other classrooms. The effect of school life on the anticipations and self-comparisons of teachers and students is reciprocal and mutually reinforcing; the people in each individual classroom are affected by the people in all of the other classrooms. Gradually, the same issues come to hold the attention of people in every classroom, and whenever students and teachers engage in anticipation or comparison their attention is pulled to the same, shared set of school issues.

In this chapter, I am suggesting more than the idea that individuals in a school are influenced by the people they find themselves surrounded by; they are influenced in a particular way, so that a special kind of interaction comes to exist in a school: coordination of behavior in different classrooms. We are all constantly influenced by events in our lives--the weather, angry or friendly treatment by someone, being told of a job opening--and we adjust to these events. But sometimes the events that influence us draw us into a special self-sustaining and reciprocal relationship with other people; the relationship itself plays a much longer role in each person's life than do the actions of any of the people in the group. Markets are an excellent example of such a relationship: farmers selling wheat are a part of a relationship that sets prices, determines earnings, creates strong incentives for future decisions about how much to plant (and organizes all of the wheat farmers in a market relationship that becomes a major force in their lives). Teachers' and students' adjustments to the actions of their colleagues and schoolmates pull them into just such a relationship. It is not a market relationship; people in schools are not primarily engaged in exchanges of well-defined goods and services. But teachers and students have this in common with farmers: they are part of a broad social relationship that provides information vital to their personal choices and actions, and the relationship of which they are a contributing part becomes a central organizing force in their lives. For teachers and students, life in school produces a poweful by-product: the coordination of behavior in all of the classrooms of a school.

The examples discussed in this chapter are intended to provide some initial support for the suggestion that one central, and little-remarked, consequence of the interaction of people in a school is to create a rough, unofficial coordination among all the classrooms of each school. I now turn to the question of the significance that this special kind of informal coordination has for our understanding of how schools work.

Perhaps what happens at the beginning of a new school year can serve as an example of how school life influences a classroom and attunes it to other classrooms. While the struggles that take place within each newly constituted classroom determine how the people in that classroom will govern themselves and how they will treat each other, there are a number of pressing questions that classroom struggles cannot answer. The answers to these questions come from outside the classroom--from the other classrooms, and from the whole school. How prepared are teachers and students for each other, and for their school tasks? What should they expect of themselves and each other--what kind of work, what kind of achievement, what level of emotional development, what problems? What standards should they apply to themselves, and how does their future depend on the standards they choose? The self-interest of each teacher and each student presses for answers to these questions, questions that orient them to the context that all of the teachers and students in the school share and that all have made. What single classrooms cannot provide--and what classrooms tied together by proximity, by the sequence of classes, and by the relationship of a shared future do provide--is information on which to base clear, direct comparisons and anticipations.

As students and teachers begin a new school year, they use school comparisons to clarify their levels of preparedness and the standards to which they expect to be held. Because their comparisons are grounded in the experiences of people in all of the other classrooms in the school, the issues that frame each person's choices are shared issues. Teachers compare the preparedness of their new students for the new year's math book; students compare the strictness and the demands of their teachers with their friends'. The orienting, pointing-out effect of anticipation works similarly: as people tell each other what to expect of the difficult students or the prickly teachers, and what will happen if they don't conform to what is expected of them, they clarify and standardize the choices faced by each teacher and student as they engage in teaching and learning.

So the life of the school draws teachers and students into a tightening relationship with each other, a relationship that adjusts each of them to the behavior of the others. The essential standards and the expectations of teachers and students are not imposed by the authority of principals or curriculum experts. Instead, these standards and expectations are the product of the experiences of self-interested students and teachers, who are themselves information-collectors, comparers, and anticipators. The compelling force of the shared school experience is its unavoidable impact on the lives of teachers and students. They must live with each other now and in classes to come; their choices are inevitably interlocked; and the interlocked choices coordinate them by adjusting their behavior to that of the other teachers and students in a school.

Bibliographical Note for Essay 3

For broad analyses of coordination through interaction, see Lindblom, Politics and Markets (Basic Books, 1977); and The Intelligence of Democracy (Free Press, 1965). For an economist's view of the internal coordination of organizations, see Williamson, Markets and Hierarchies (Free Press, 1975). The experiential literature of schooling, critiqued by Averch et al., How Effective Is Schooling? (Educational Technology Publications, 1974), often implicitly suggests that schools tend to move toward consistent, coordinated behavior patterns. References to school tracking and ability grouping can be found in Rosenbaum, Making Inequality (Wiley, 1976).

Essays on School and Policy Analysis

Essay 4: What Difference Do Classroom Interactions Make? Teachers, Students and Reciprocal Sovereignty

by

Edward W. Pauly

Institution for Social and Policy Studies
Yale University

1982

This research was supported by the National Institute of Education under grant NIE-G-79-0084. NIE is not responsible for the content of the research.

There are two conventional answers to the question of how students' cognitive achievement is determined. One emphasizes the characteristics of students; the other, characteristics of schools. Both answers have had great impact on the way we, as a society, think about education and education policy. Since this book proposes a very different approach to thinking about educational problems, it may be useful to outline briefly the orthodox versions of school policy thinking.

One major strand of educational thinking, usually associated with the Coleman Report, holds that students' backgrounds determine their cognitive achievement. According to this view, school-to-school differences don't matter much; what does matter is the collection of attributes that are already fixed when students come to school. These include students' IQ, membership in a particular part of the culture, and what are held to be home- and community-determined factors such as motivation, aspirations, family support, and social class. These characteristics of students are said to be the controlling factors in their educational experiences and cognitive achievement, and schools are believed to play only a very modest role in affecting cognitive achievement.

The other established approach to educational thinking holds that characteristics of each student's school determines cognitive achievement. Classroom attributes such as the amount of time spent on reading or mathematics tasks, the use of particular instructional methods, and the teacher's skill in classroom management and instruction are seen as the primary influences on what students learn.

The policy consequences of both approaches have been substantial--and puzzling. Government agencies and local school systems searching for ways to improve student achievement have invested heavily in policies associated with each approach.

The 'student characteristics' view has led to policies aimed at improving students' home resources, and policies such as busing that alter the mix of peers to whom students are exposed. Reviewers have found that these policies have had distinctly mixed effects; moreover, studies demonstrating the frequency of unexpectedly high or low achievement for students, given their home backgrounds, raise puzzling questions about the degree to which student characteristics really dominate what students learn.

The 'school inputs' view has led to the broad sweep of school aid, classroom intervention, and instructional improvement policies of the 1960s and 1970s. Research has repeatedly found that the trouble with these policies has been their inconsistent results and the persistent difficulty of replicating apparently 'successful' projects: what works in one time and place doesn't in another.

At the end of this chapter, and in Part II of the book, I will return to the evidence on these approaches. What the preceding chapters have led up to, and what I will now present, is an alternative to the orthodox views of how schooling works, a view that has significant consequences for the ways that we think about educational policy.

This book disputes the conventional answers to the question of how schools work. Thus far I have argued, in a number of different ways, that it is not so much school inputs or student characteristics that make schools work as they do; it is the daily interactions among teachers and students.

But in order to undertake a systematic and clear account of teaching and learning that puts classroom interactions at the center of policy makers' attention, we need more than a general awareness that interactions among teachers and students matter. We need a specific focusing device for our ideas about classroom interactions--a conceptual lens through which to look at the workings of schools.

The preceding chapters indicate what kind of conceptual lens is called for. The lens should focus our attention on the purposive behavior and political struggles that take place in classrooms and in schools, and on the striking differences in how people in similar schools and classrooms carry out identical policies. And the conceptual lens should focus our attention on the relationships that stimulate or inhibit effective teaching and learning in schools.

I. Reciprocal sovereignty

The name that I have given to the conceptual lens is reciprocal sovereignty. I want to suggest that reciprocal sovereignty captures the distinctive and powerful characteristics of day-to-day school interactions--

things that conventional approaches to the analysis of education have missed. To begin, I will explain where this new phrase comes from.

"Sovereignty" refers to the power of each person in the classroom, student as well as teacher, to command the attention, resources and responses of other people in the classroom. Teachers and students use their sovereignty to get what they want from others; in particular, they use it to obtain desired behaviors from others. This sovereignty can be exercised in many quite different ways, and it can be used to accomplish a wide variety of ends. Classroom sovereignty rests partly on commands and controls that are voluntarily accepted by fellow classroom members. It also rests on manipulation and coercion. As teachers and students approach their most intensely held purposes, they are increasingly likely to enforce their will on the people around them--whether by a student's disruption of the classroom, or a teacher's arbitrary decision. Sovereignty in classrooms involves the potential of threats, as well as voluntary acquiescence to another person's demands.

Since each person in a classroom has, and exercises, sovereignty over those around him or her, this sovereignty has a two-way property: each teacher and student is partly controller and partly controlled. And since each person recognizes that the others are also sovereign in important ways, people in classrooms are acutely aware of the essential give-and-take quality of classroom sovereignty. Reciprocal sovereignty has the crucial characteristic of recognized interdependence among many powerful controllers in each classroom. Teachers and students take account of both their sovereignty and their reciprocal relationships when they act.

"Reciprocity" thus refers to the fact that in classrooms, each student and teacher is a powerful supplier of behaviors wanted and needed by the others. People give these needed behaviors to others with the knowledge that they will soon need return behavior from the same people who are now, temporarily, their beneficiaries. The symmetry of this relationship--I need things from you, and you need things from me--builds reciprocity into the classroom.

Drawing together these ideas, we can define reciprocal sovereignty as follows: Reciprocal sovereignty is the process of organizing the joint exercise of power among people who are both controllers of, and controlled by, each others' actions.

Reciprocal sovereignty grows out of some attributes of classrooms that I have often cited in the preceding chapters; these make up the context in which reciprocal sovereignty takes root.

- Classrooms are small, densely populated, and enclosed places. Within their classrooms, people are highly exposed and vulnerable to each other. As a result, the actions of each individual student and teacher take on more potential to hurt or to help others than they would in other settings. In classrooms, the power and significance of an individual's acts become magnified.
- People in classrooms are very concerned about how they will be treated by the other people there. Power relationships are therefore at the center of everyone's attention; how people treat each other is never a trivial issue.
- The ability to command other people's attention, resources, and responses belongs to everyone in the classroom. Conventional analyses of schooling tend to treat teachers and students very differently--

> teachers as active, intelligent, powerful implementers of policy, students as submissive recipients of policy. The reciprocal sovereignty lens regards teachers and students (and students from low income families as well as students from affluent families) as participants in the same process, and as responding to the same issues in their experiences.

Looking at schools through the lens of reciprocal sovereignty differs from conventional ways of looking at schooling in several ways. First, in important ways, the experiences of people in classrooms are seen as uncontrollable and highly variable; this produces classroom behavior that is tailored to conditions of uncertainty. Second, the school setting is seen differently: dynamically, as an evolving and time-defined environment. Third, the primary decisionmaking process of schooling is seen as a decentralized process of self-regulation carried out by the people in classrooms, who "solve" their own "problems."

Uncertainty and the classroom's self-protectiveness. The reciprocal sovereignty lens reveals the extent to which school behavior is formed in reaction to the deeply-felt uncertainties of school life. Teachers and students are subject to each others' control, and they know that this is so. They know that the demands they will face from others are impossible to foresee, and therefore that uncertainty about how they will be treated is built into their predicament. Teachers also know that how much their students learn, and when and how they learn it, is sure to swing wildly from year to year, and even from week to week. They also know that their ability to gain acknowledgments of their own helpfulness from students will not last, always remaining uncertain and beyond their control. Students know that the demands and pressures they receive from peers and teachers are always in flux. Because the workings of reciprocal sovereignty are unpredictable, they breed uncertainty and anxiety. The anxiety of not knowing how one will be treated is part of classroom life.

Teachers and students act to protect themselves against the anxieties of the classroom by using the mechanisms of reciprocal sovereignty. The same mechanism that creates their predicament provides the solution to that predicament. Teachers and students use reciprocal sovereignty to gain reassurances about how they will be treated. These reassurances take the form of concessions granted to each other--assurances that it is all right to move ahead in the lesson, that a wrong answer will not be cause for humiliation, that the rules of a competition are stable and fair. The reciprocal exchange of commands provides the security needed for each person in the classroom to accept the vulnerabilities of teaching and learning. Reciprocal sovereignty is used to map out for each person what is "safe" to do in the classroom.

The net effects of people's responses to the uncertainties and anxieties of the classroom is that classroom interactions serve a function of self-protection in classrooms. Even an offer to help another person may be stimulated by the knowledge that helping is self-protective: it evokes a reciprocal favor. Helping behaviors are rarely thought of as exercises of sovereignty, but when they function as conditional offers of support, they strengthen and protect the helper. Teachers and students uncertain about how others will treat them will quite sensibly hesitate to commit themselves, or invest their energies, in classroom learning activities--until they can rely on the exercise of reciprocal sovereignty to regularize and clarify the behavior of others toward them. The choices that students and teachers make--

choices of how energetically to study math, when to move to the next task, how to respond to negative evaluations of one's work--depend on how well they are able to reassure and protect themselves through the use of reciprocal sovereignty.

Time and sustained classroom life. The passing of time is also tightly bound up with reciprocal sovereignty. It is the repeated exercise of reciprocal sovereignty, over time, that builds up layers of demands, responses, and expectations, and thereby provides the foundation of information and agreements that make sustained classroom life possible. The exercise of reciprocal sovereignty slowly transforms strangers into people highly sensitized to each others' behavior--with all the possibilities for support and exploitation that inevitably result.

For teachers and students to cover a body of instructional material successfully, they need to develop cooperative relationships and channels of communication that will allow them to sustain a working relationship. That relationship can only be built up slowly, over time. On the other hand, in classrooms in which people are caught in a pattern of sustained failure, the layers of hostile and distrustful relationships built up among teachers and students that keep people failing may simply not be visible if one takes a one-time, "snapshot" perspective. These problems cannot be solved by policies that focus exclusively on teachers or on instructional programs, or on students, instead of on the interdependence, and the reciprocal sovereignty, of teachers and students.

The relatively long duration of the school year, and the relatively fixed assignments of teachers and students to classroom groups, make drastic demands on people's ability to get along with each other, to sustain working relationships, and to develop predictable sequences for solving classroom problems. Without a mechanism that allows for the gradual acquisition and evolution of classroom rules, skills, techniques, and a broad information base about the people in each classroom, little would be accomplished.

Decentralized self-regulation in schools. The conceptual lens of reciprocal sovereignty focuses our attention on the strikingly large amount of school activity that goes into adjusting conflicting wants, balancing contending values against each other, and sorting out the problems created by the presence of a collection of energetic, purposive teachers and students inside a single set of walls. By this point in my argument, two important features of reciprocal sovereignty have become apparent: the decentralized nature of its operation, and its capacity to regulate teachers' and students' classroom actions. For analysts and managers whose habit is to regard decision making as a centralized, hierarchical process, these features are startling.

The exercise of reciprocal sovereignty means that the conflicting behaviors and conflicting concerns of people in a classroom get adjusted to each other. In practice, this is what reciprocal sovereignty looks like: teachers and students stop and start each other; they schedule each other; they qualify and amend each others' plans and intentions; they tell each other to try again, to change direction, to shift attention, to give aid, to resolve a conflict. When teachers and students disagree over who should do what, processes of reciprocal sovereignty produce the resolution. Reciprocal sovereignty resolves issues by altering each conflicting person's demands enough so that relatively predictable and organized behavior can go forward.

Reciprocal sovereignty is especially well-suited to the problem of resolving classroom issues that require balancing disparate values and actions. Teaching and learning depend not just on instructional practices but also on individual preferences and values, emotional ties among classroom members, and social and competitive relationships. Adjusting this complicated web of relationships to some new pressure or demand, whether from inside the classroom or not, is a task that greatly exceeds the cognitive capacity of even the best managers or analysts. But the workings of reciprocal sovereignty reveal the differing value weights that each person in the classroom places on the object of the pressures, and communicate this crucial information to all participants. Teachers' and students' *actions* show how much importance each one attributes to each complicating factor. Reciprocal sovereignty provides an effective substitute for the impractical effort to calculate how to adjust classroom behavior to each now pressure; it determines *through interactions* who will prevail, in what limited ways, and what reciprocal concessions they will make to others in the classroom, when adjustments in behavior are required.

Reciprocal sovereignty adjusts the conflicting demands of teachers and students; it resolves highly complex questions of how to balance disparate values in teaching and learning. In these senses, we can say that important parts of classroom life are regulated by the people in each individual classroom, acting to solve their own problems for themselves. We may not like the results that they achieve, but we must acknowledge that they are engaging in a kind of self-regulation, and they are doing so in a distinctly decentralized way. Central authorities, legal principles and administrative rules can all be subverted--in most cases, quite easily--by the exercise of reciprocal sovereignty. So it appears that the really significant source of regulatory power over classroom behavior is held by the people in each classroom.

An important consequence of the decentralized self-regulation of classrooms is that the same disagreement, or the same external pressure, will be resolved differently in different classrooms--even within the same school. Whether or not this characteristic process of resolving conflicts in classrooms is in some sense preferable to some alternative decision making process may be of little consequence. Reciprocal sovereignty is what actually goes on.

Two further implications of reciprocal sovereignty should be recognized. According to the line of analysis I am proposing, instruction is not controlled by the teacher (or policy maker); the control of instruction is dispersed among the people in each classroom, and it is dispersed and organized differently in each classroom. As a consequence, a wide range of instructional activities will be, in effect, controlled by reciprocal sovereignty: the allocation of teachers' and students' time, the choice of topics of instruction, and the availability of help, cooperation, and support (for both students and teachers), for example. Proponents of the 'school inputs' view have tended to see these activities as controllable by policies; and in fact, the school improvement efforts of the 1960s and 1970s were based on the assumption that predictable changes in school and classroom behavior could be produced by careful policy interventions. The reciprocal sovereignty view disputes this assumption on the grounds that classroom attributes are controlled by the people in classrooms; only by influencing their independent actions can policies affect what goes on in schools. Policies may affect people in classrooms, but--because of the workings of reciprocal sovereignty--not in consistent, predictable ways. Thus, it follows that school inputs determine neither the variation among classrooms in teaching and learning activities, nor the impact of teaching and learning on behavior and achievement. (Empirical support for this proposition will be discussed later in this chapter.)

The second implication is similar. The regulatory effects of reciprocal sovereignty extend to all forms of classroom behavior, by teachers and students alike; and all classroom behaviors are susceptible to being manipulated and altered by the workings of reciprocal sovereignty. Consequently, reciprocal sovereignty is capable of altering anti-social behavior and behavior that inhibits teaching and learning. Analysts in the 'student characteristics' tradition have regarded these student attributes (and related background variables) as fixed, and beyond the influence of school experiences; and they assume that student achievement is controlled by student background. In contrast, the reciprocal sovereignty approach argues that important and effective teaching and learning activities are controlled by the people in schools and classrooms. Consequently, the arguments developed here suggest that student characteristics determine neither classroom variation in teaching and learning, nor the impact of teaching and learning on behavior and achievement. (Empirical support appears later in the chapter.)

II. The effects of reciprocal sovereignty

How does the exercise of reciprocal sovereignty affect student achievement? Conventional views of schooling typically attribute achievement gains to school inputs or to student characteristics. The concept of reciprocal sovereignty suggests that these approaches, in their search for the "right" conditions for learning, may miss the classroom processes through which teachers and students determine for themselves how they will teach and learn. When we focus on the operation of reciprocal sovereignty in classrooms, we see a number of powerful educational processes taking place:

The allocation of resources within classrooms. Reciprocal sovereignty determines the uses of time in classrooms, and determines the availability of responses, of support, and of interaction that are necessary for classroom achievement. Students' demands induce teachers to search for additional ways of explaining a problem; they allocate the teacher's time to certain students. Teachers' demands induce students to put their energies into projects requiring individual initiative; they stimulate students to use peer pressure to control disruptive classmates. Among the crucial educational allocations produced by reciprocal sovereignty is each classroom's determination of students' and teachers' time "on task."

The adaptation of teaching and learning to particular classrooms. Reciprocal sovereignty induces teachers and students to tailor their teaching and learning activities to the idiosyncracies of the individuals in each classroom, and to the special patterns of interaction that have been established in that classroom. This process of adaptation takes place because people in classrooms recognize that they are dependent on the people around them for things that they need; their dependency induces them to become closely attuned to the highly personal attributes that are communicated by other people's demands. In the course of exercising reciprocal sovereignty over one another, students and teachers exchange information: about their fears and limitations, about where their attention lies, about their interest in a lesson (this applies to teachers as well as students!), and about the connections between classroom tasks and their personal agenda and values. Because this information identifies the intensely-held individual needs and classroom social arrangements for which standard curricula and methods most need to be altered and adapted, and because it is obtainable from no other source but classroom interaction, it is vital for classroom teaching and learning. Reciprocal sovereignty adapts achievement-related tasks to the needs and idiosyncracies of the unique collection of individuals present in each classroom.

The identification of opportunities for cooperative action. The back-and-forth exchanges of reciprocal sovereignty are a necessary condition for students and teachers to discover, and voluntarily enter into, opportunities for joint, cooperative action with others in the classroom. In order for this to happen, there must be more than a simple exchange of personal information. Evaluations, preferences, and interpersonal judgments of character and competence are essential ingredients when teachers and students agree to commit themselves to cooperative action. The normal workings of reciprocal sovereignty thoroughly air and reveal these judgments and wants. Among the educational results of cooperative classroom action are teachers' and students' acceptance of status and rank groupings, and their use of competition as a part of teaching and learning.

The creation of channels of communication and interaction based on emotional bonds. The reciprocal exchange of demands and controls creates a broad range of emotional bonds among teachers and students. These include bonds of obligation, trust, and gratitude, and also bonds of intimidation, humiliation, ridicule, dependency and exclusiveness. These emotional connections can be destructive of various teaching and learning activities, or supportive of them. They share, however, a single educational effect: they create a rich fabric of possibilities for communicating ideas and perceptions about teaching and learning activities. Emotional bonds can therefore function as the communications grid across which are transmitted the signals of teaching and learning.

* * *

One of the central puzzles facing analysts and policy makers in public education for the last fifteen years has been the inconsistent results of school policy interventions and experiments. The leading analysis of the 'school inputs' approach, now ten years old but still accurate, finds simply, "Research has not identified a variant of the existing system [of schooling] that is consistently related to students' educational outcomes." (Averch et al., 1972) In other words, the conventional view that sees alterations in school inputs as producing predictable changes in students' achievement has not been found to be empirically supportable. In a careful analysis of the quantitative studies bearing on the 'student characteristics' approach, Murnane concluded that "the most notable finding is that there are significant differences in the amount of learning taking place in different schools and in different classrooms within the same school, even among inner city schools, and *even after taking into account the skills and backgrounds that children bring to school*." (Murnane, 1981; emphasis added.) That is, the conventional view that sees students' characteristics as the controlling factor in their achievement does not hold up.

Despite extensive and increasingly sophisticated studies by social scientists, and despite the best managerial efforts of educators and administrators, the conventional conceptual apparatus seems not to offer a very powerful or compelling understanding of how schools work or how they are affected by policies.

The conceptual framework presented in this book attempts to suggest how we might make sense of the policy problems involved in improving student achievement. To the extent that the conventional views of schools have led us on a search for sure-fire "treatments," whether they involve better school inputs or efforts to alter student background characteristics (or to use student characteristics as a basis for assignment to a school), we may have been misled. Of course, it is very difficult for any of us to develop a tolerance for not having answers that tell us what to do. But the line of argument of this book suggests that traditional answers prescribing fixed policies that seek predictable school change will not work. The reason is simple: the actions of people in schools substantially transform all policies aimed at influencing their behavior. It is frustrating that the two conventional approaches don't provide the reliable answers that we wish for; it suggests that there is nothing to be done. But there are things to be done; they are, however, more difficult and more subtle than are the conventional prescriptions.

It seems to me that the consistent theme of both the Averch and Murnane reviews is the extent to which variations in teaching and learning are produced by the people in individual schools and classrooms, interacting with each other and working out arrangements among themselves about how things will be done in their classrooms. Instead of a hard and firm answer to the question of how schools work, we have come to focus on very diverse, decentralized human processes that determine how education is conducted: the making of the classroom constitution (chapter 1), the coordination of individuals' activities within schools (chapter 3), and the exercise of reciprocal sovereignty. It is the always-changing process of interactive problem solving, rather than the hard answers provided by school inputs and student characteristics (important as they may be), that seems to account for the research findings of Averch and Murnane.

The question for policy makers, then, is not what the effect of a given policy will be; it is, instead, how will each classroom adapt and transform the policy? Since reciprocal sovereignty works differently in every classroom, we must expect the process of policy response to be so particularized as to defeat any synoptic, strictly imposed educational "treatment." This does not mean that policies cannot contribute to improved teaching and learning; it does mean that policy makers have neither the analytic capacity to prescribe any specific improvement, nor the power to enforce it.

But if we must learn to tolerate the absence of hard answers about the specific effects of particular policies, what can we put in the place of old ideas about school inputs and student characteristics? The thesis of this book is that if we put classrooms, teachers and students at the center of our analysis, and consider the processes through which they work to solve their own problems, we can act in ways that enhance the energies and initiatives that they bring to schooling. As the succeeding chapters will show, some policies support the evolution of effective classroom rules and others do not; some policies induce people to discover new information about their settings and others do not; some policies create incentives for behavioral responses that foster cooperation and others do not. These examples suggest that school policies may have their greatest effects through the independent, decentralized classroom responses they evoke, and through the adaptability over time of policies to changing classroom responses.

Bibliographical Note for Essay 4

The "student characteristics" literature begins with Coleman et al., Equality of Educational Opportunity (Government Printing Office, 1966). The "school characteristics" literature is criticized in Averch et al., How Effective is Schooling? (Educational Technology Publications, 1974), and in Berman and McLaughlin, Federal Programs Supporting Educational Change, Vol. I: A Model of Educational Change (The Rand Corporation, 1974).

Discussions of uncertainty in classrooms are found in Sarason, The Culture of the School and the Problem of Change (Allyn and Bacon, 1971), and Jackson, Life in Classrooms (Holt, Rinehart, Winston, 1968). The significance of the passage of time for classroom interactions is implicit in much of the experiential literature, described in Averch et al. The possibilities of decentralized social problem solving are described in Lindblom, Politics and Markets (Basic Books, 1977).

Citations on page 15 of this essay are to Averch et al., and to Murnane, "Interpreting the Evidence on School Effectiveness" (Teachers College Record, Fall 1981). References to research on adaptation in schools, on pages 15-16 of this essay, are based on Berman, McLaughlin, and others, Federal Programs Supporting Educational Change, especially volumes I, II, III, VII and VIII (The Rand Corporation, 1974-1978).

Essays on School and Policy Analysis

Essay 5: Analysis When Conventional Analysis Won't Work: School Officials and the Uses of Interactive Analysis

by

Edward W. Pauly

Institution for Social and Policy Studies
Yale University

1982

This research was supported by the National Institute of Education under grant NIE-G-79-0084. NIE is not responsible for the content of the research.

All school officials, from principals to superintendents, constantly engage in analysis as a means of deciding what to do. Whether it takes the form of a formal evaluation of a reading curriculum, or an informal visit to a few classrooms in response to questions from parents, analysis is a central part of their job. In this chapter, I will argue that the conventional descriptions of how analysis is done, and how it *should* be done, are mistaken, inappropriate and misleading when applied to schools. The reasons stem from the school attributes described in the preceding chapters, and particularly from the effects of reciprocal sovereignty in classrooms. Conventional analysis won't work in schools, at least in the ways that we expect it to work; it won't tell us which kinds of teaching and learning are effective and which ones aren't; and therefore it won't tell us what policies and programs the schools should choose.

Yet analyses and evaluations are still powerful and useful for school people--although not for the conventional reasons. First, the enterprise of analysis often prompts both officials and teachers to think about their work in new ways. Second, while conventional analysis doesn't work in schools, other forms of analysis do. I will argue that people in schools come to rely on what I will call interactive analysis, which emphasizes the particular kinds of information and knowledge that are available and productive in the day-to-day world of schools.

I. The preconditions for analysis and evaluation

People in schools engage in analysis all the time; but much of what they do is regarded by critics as insufficiently technical or "scientific" to be taken seriously. As a result, the labels "analysis" and "evaluation" have come to be reserved for what is actually a small, specialized corner of the broad territory of serious analytic thinking. This chapter will argue that the analytic enterprise is carried out in a variety of potentially powerful ways, many of which we fail to appreciate in our concentration on the search for scientific, and thus presumably "authoritative," analyses and evaluations. But the record of attempts at "authoritative" evaluations has been disappointing. In order to find out why this is the case, we need to focus on the prerequisites that are necessary for the success of the technical model of analysis--the approach that has become the accepted, esteemed, conventional approach.

Conventional analysis is a subcategory of what I shall call analysis, which comprises the detailed examination of the effectiveness of school and classroom activities in fostering student achievement, conducted so as to clarify the choices and constraints facing school people. The conventional approach to analysis and evaluation examines school effectiveness in a particular way: it identifies the differences in average student performance that are associated with specific, systematic differences in school and classroom practices; and it does this in a way that aims to make predictions about the consequences for student achievement of a range of alternative choices. In other words, the analyst seeks to discover "policy variables" that will lead to high achievement in schools and classrooms that adopt the recommended policies. This narrow and technical approach is consistent with the (much more inclusive) notion of evaluation with which we began; but the conventional model must satisfy three preconditions if it is to work.

First, the information to be used in the analysis must be available and be reasonably accurate. Second, the information myst be systematic, standardized and comparable--it must reflect meaningful differences among schools and classrooms. Third, the information must reflect conditions that are stable over time.

The efforts of school officials and researchers to fulfill these preconditions have absorbed huge amounts of their energy, and a large literature has grown up to instruct and counsel the would-be educational evaluators. People have tried hard, very hard, to make the conventional model of analysis work in schools. It is all the more striking, then, that all this labor has produced little more than an essentially self-contradictory, disappointing collection of studies, vast in number but highly limited in what they can tell us. The reason for this is that none of the three preconditions for the conventional model of analysis is even remotely consistent with the reality faced by people in schools.

In schools and classrooms, the dominant reality is the phenomenon of reciprocal sovereignty. Because teachers and students engage in intense and private efforts to govern the ways they will treat each other in the classroom, official demands for consistent procedures and policies are subordinated to the interactions of the people in that classroom. Officials are peripheral to the classroom's world, a world that is more the product of reciprocal sovereignty than of an official action.

When officials try to collect information about teachers' instructional behavior and students' achievement, they discover that the availability of that information is effectively controlled by teachers and students. Access to classroom information is typically conditional: people in classrooms cooperate with information-gathering officials if there will be no negative consequences of doing so; otherwise, they present themselves cloaked in the guise of compliance with official policy, behind which they go on as before. This version of classroom resistance to evaluation is often seen by officials as dishonesty or sabotage, and from the top-down chain-of-command perspective of conventional analysis it is hard to avoid such a conclusion. But the logic of classroom life is not built on defiance. People in classrooms want to develop and live the shared life that they have worked among themselves, without hassles or implied threats. This may be self-centered; that is the source of the classroom's energy and creativity. It is not obstructionist. But the implied criticism and the potential threat of enforced change can trigger the mixture of vulnerability and self-protectiveness that undermines the collection of accurate, complete information. Since people in each classroom can edit their behavior when they are observed, their selective revelations present a highly distorted picture to observers.

The second precondition for evaluation is that its data must be systematic, standardized and fully comparable across classrooms. This problem has two parts. First, it is obvious that without comparable outcome data, the analyst cannot make comparisons to discover where schooling is effective and where it is ineffective. But the second difficulty, obtaining comparable input (or "process") data, is even more disabling. Without systematic, comparable measures of classroom behavior, the analyst cannot even determine what instructional practices are being used in a particular school or classroom. In order to evaluate a practice and trace its consequences, one must be able to define it, identify it, and compare it. But if classrooms vary in how they carry out a given instructional program--and if their variations are part of the distinctive give-and-take of reciprocal sovereignty that adapts *every* program to the people in each classroom--officials doing evaluations will find more evidence of variation than of concrete, identifiable programs. The confusing differences among classrooms may focus evaluators' eyes on whatever sign of a standard program or practice they can find, even if its presence is mostly an optical illusion produced by a history of using a single name to refer to what in reality is a wide variety of reading programs. When Jeanne Chall studied the strict phonics method for teaching beginning reading in 1967, she visited schools where it was ostensibly in use--but the officials "who accompanied me were quite embarrassed to find that teachers were doing many things that were 'not permitted' by the method" (p. 284). Her conclusion was not that some demonic non-compliance was at work, but that *no* method can avoid the unconscious "improvements" that creep into every classroom's response to a given standard. The reality of school officials' experience is that reciprocal sovereignty introduces so much variation into teaching and learning as to transform them into quite separate and distinct enterprises in each classroom. Officials who try to collect information for an analysis will thus come up with information that is impossible to systematize, non-comparable across classrooms, and so fragmented that it constantly diverges from any shared standard.

The third precondition for analysis is that it must be based on school and classroom information that remains stable over time--that is, today's description of a classroom's teaching and learning processes is connected in a reliable, predictable way with the teaching and learning processes that will exist next week in that classroom. School officials face the problem of how to observe and analyze an ongoing process, and again they discover that the information they can collect will not support the weight of the assumptions built into conventional analysis. There are two reasons for this. First, the relationships that produce teaching and learning in a classroom grow and change. The events of October are built on the exchanges that took place in September, as people in classrooms adapt and continue what has seemed to work and change what has not. These changes in a classroom are hard to track, because the options and choices available in January do not correspond to the ones that faced people in October. Second, the classroom's inhabitants actively respond to pressures on their habitat--even including the pressures of an evaluation, which clearly signals that some aspect of their lives is being watched and could probably use some sharpening up. So the classroom changes itself in response to outsiders as well as insiders. Both of these sources of instability and change in the classroom make it nearly impossible for school officials to trust their observations to be lasting, or even revealing of anything other than a collection of soon-to-be-modified, momentary behaviors. It is confusing, at best, to try to organize and use such unstable information for policy purposes. As a consequence, officials inevitably discover that their data are highly perishable, and poorly suited to an analysis that aims to prescribe programs and policies for the future.

If the experiences of school officials are so inconsistent with the assumptions that lie behind standard approaches to analysis and evaluation, we should expect to see evidence of these problems in the documentation of past analyses. As for the openness of people in schools to evaluation, Milbrey McLaughlin's survey of the various evaluations of the federal government's massive Title I program for disadvantaged students found that officials encountered such broad resistance to data collection, and resistance of so many kinds, that access problems, rather than analytical ones, dominated evaluation findings. Usable information was nearly impossible to obtain; solid inferences, completely so. Her finding was that many of the Title I project reports "resemble educational travel brochures, with extensive anecdotes and little objective data to support claims of 'success.' The evidence presented is often unrepresentative, impressionistic, and incomplete, if not false (p. 22) . . . (For one large, carefully conducted national survey) most of the completed forms lacked achievement data . . . despite the energy and effort invested in the striking of consensus and alleviating concern, many local schoolmen still hesitated to furnish the achievement data USOE requested, and others did not collect any standard information at all" (pp. 53-54).

The evaluation literature provides a catalogue of findings that have much more in common with our descriptions of program variations created by reciprocal sovereignty than with the crisp results and clear comparisons that analysis is supposed to produce. The Rand Corporation's massive 1972 review of the literature found that "research has not identified a variant of the existing system that is consistently related to students' educational outcomes. The term 'a variant of the existing system' is used to describe the broad range of alternative educational practices that have been reviewed . . . . We specifically include changes in school resources, processes, organizations, and aggregate levels of funding" (pp. 171-172).

The results of the studies were simply, and devastatingly, inconsistent with each other. A careful study of model programs (called "treatments") developed for use in the Head Start early schooling program found that "classes under the same treatment label have differing experiences (p. 118) . . . participants, not surprisingly, tend to be unclear about what they are to do or why they are to do it. Also, no two situations are alike" (p. 125). And this was an experiment in which participants were extensively trained and supervised to comply with an instructional model. How could anyone compare such changeable, impossible-to-label classrooms?

The pattern of 'no pattern' does have a certain bizarre symmetry to it. The formlessness of the evaluation literature should not be taken as evidence for the hypothesis that it does not matter what educational approach one uses. It is, rather, a product of a consistent pattern of contradiction, variation, and divergence. As the Rand team found, "The literature contains numerous examples of educational practices that seem to have significantly affected students' outcomes. The problem is that there are invariably other studies, similar in approach and method, that find the same educational approach to be ineffective. And we have no clear idea of why a practice that seems to be effective in one case is apparently ineffective in another" (p. 172). There are effective classrooms; but conventional analyses have not been able to pin down the sources of their effectiveness.

That fact is discouraging. But the more sanguine analysts have found in it some basis for hope and speculation. The Rand team took the school analyses as lessons in how to improve research, and suggested that the lack of clear evidence in support of costly, risky programs could be taken as an argument against continued funding. Chall and others write approvingly of the healthy practical adaptations and modifications of programs, despite the fact that these adaptations effectively ruin any formal, conventional evaluation. And Richard Murnane's 1981 survey of evaluations sees in "human resources" (and human incentives) the key to why the same policy turns out differently in each setting. All are thoughtful responses that make the best of a troubling situation.

And why is it so troubling? Simply put, because the preconditions for evaluation are not met, and appear never to have been met. The accurate, available information that can serve as comparable descriptions of school practices that are stable over time simply cannot be found in the evaluation literature.

But evaluation and analyses go on anyway. Despite the extraordinary problems faced by every school official involved in analysis, and despite the depressingly familiar, repeated pattern of positive findings followed by negative ones, officials behave as if evaluation were not only a normal and appropriate enterprise for them, but an engaging and useful one. Their experience of its usefulness for policy making to the contrary, they continue to participate in analysis and evaluation. Teachers tend to be less enthusiastic about participating in evaluations. Yet they, too, perhaps defensively and perhaps less than candidly, continue to put up with these unproven instruments of consistently inconsistent analysis.

II. The side effects of evaluation and analysis

The paradox of conventional evaluations is that they cannot provide the information base for the policy choices that are their official justification for existence; yet they continue to be used, by people who are neither foolish nor likely to engage in bureaucratic plots. I will suggest that even without any intelligible, usable results of a conventional analysis--even without any results at all--these events are the occasion for people in schools to think about teaching and learning in distinctive and powerful ways. It is the interactions, the give-and-take among school people, that are the crucial side effects of an evaluation; these interactions stimulate new ways of thinking and analyzing among officials and teachers. Whatever the accuracy of the technical justifications and assumptions for evaluations, I will suggest that school people continue to engage in them for a little-noticed reason that the politics of evaluation turns out to be a powerful device for stimulating broad problem-solving efforts and powerful confrontations. Conventional analyses may be too weak to base policy choices on; but their strength in promoting the give-and-take that stirs up the policy process (and the thinking of the people in that process) is as great as it is little noticed.

The transformation of an analysis into a set of political perceptions comes about because of the ways evaluations are carried out in the grass-roots world of schools. Evaluations call for a variety of concrete actions: the school programs that are to be assessed must be clearly identified; educational goals and desired outcomes must be clearly specified, and measures of these outcomes chosen; information must be systematically collected; and that information must be recorded. It is these concrete events that engage the responses, and the independent thoughts, of teachers and school officials.

Deciding to analyze a program has the unintended side effect of checking up on what activities are actually taking place in the classroom.

Are the designated math books actually being used? Were they even delivered to the classroom in the first place? And if not, is there a good reason, a disputed reason, or no reason? (Is there, in fact, any "treatment" to be evaluated?) Whether the classroom behavior that is identified as the program to be evaluated is a good idea or a bad one, it becomes the ground on which officials and school people meet. Their discourse is not a random one, and not susceptible to easy evasion, as ordinary complaints or official rhetoric can be. Marking a program for analysis means identifying it as a common property of the school--part of the school's standards. Evaluations confront school people and officials with a piece of established behavior, to examine and think about together. And as a consequence, their thinking is stimulated--and changed.

The focus on specified educational outcomes introduces a language of accountability and commitment into the transactions between officials and their subordinates. This happens when arbitrary, ad hoc and vague goals for instruction are replaced by a clear statement of how a program's official goals will be measured. The results of the sharpened and more public language of officially-sanctioned goals may produce a new consensus--for example, by preventing officials and teachers from drifting into patterns of teaching and learning that touch many areas lightly, rather than covering a few well. Or the results may reignite old conflicts between officials and teachers over the enforcement of a standard criterion on widely differing classrooms. But the simple announcement of a goal during an evaluation invokes a political process of thinking, debate, and give-and-take among school people and officials. They have something concrete to think about and to say to each other, something tied to their immediate experiences, their jobs, and their accountability and commitments to others.

These responses to newly-stated goals can arise not just from discussions of goals themselves, but from decisions about how to measure whether the goals are met. Even such a seemingly technical issue as the evaluation's instruments for measuring particular educational goals can trigger these responses. For example, the instruments in the analysis of a reading program might measure vocabulary, or paragraph comprehension, or word decoding--and the use of a particular instrument could prompt school people to think more critically and sharply about the combination of instructional activities used in their classrooms; or it could provoke disagreement on the appropriateness of a particular goal for all classrooms; or it could lead to perverse response such as teaching to the test, or shortchanging the skills not covered by the agreed-upon measures. When an evaluation is begun, officials and teachers face each other over the ground of shared vulnerability. And they think about their own actions, and each others', in new ways.

The systematic gathering of information is usually seen as a dull prerequisite for analysis. Yet what is most striking about this process is not the data that are produced (data that are badly undermined by the differences among the classrooms from which they come), but the way that the data are produced. The effect of collecting any given piece of information systematically, consistently, through the exercise of managerial power is to change school people's thinking about their relationship to each other. Information that is gathered systematically rather than casually creates expectations, comparisons, and the widespread perception of centralized management. If a principal or teacher is instructed to report on his or her school's use of federal Title I funds, no one in that school will mistake the significance of the systematic quality of what is happening: their school's information is being collected according to the same system as many other schools' information. There is an unavoidable lesson that emerges from such an experience: a dollar spent in one school is, to officials, the same as a dollar spent in another school; and both expenditures are equally demanding of justification. And no matter how sympathetic the officials doing the analysis may be, the intimate, local school experiences of scarce resources, special-category students, and involvement in a special program will also take on an official, centralized meaning. The systematic gathering of information is an occasion for people in each school to think about their tangible, official connections with the people in all of the other schools in the evaluation.

The making of a written record of evaluation information is inevitably the most controversial practical aspect of an evaluation, because the record makes the evaluation into something physical, exchangeable, and permanent. Evaluations leave a trail of paper behind them.

Information is power, and that includes information on test scores, daily attendance, the number of test-takers, subsidized lunches, substitute teachers, parents on welfare, vandalized textbooks--almost anything one can imagine. Once it has been recorded, information can be used anywhere: in an election campaign, in a newspaper's investigation of the schools, in the negotiations between the teachers' union and the school district over class size, and even in the requests made by principals or parents for their school to receive benefits that were recorded as being available at other schools. Recorded information on resources, services, eligibility for benefits, the results of school programs, the problems faced by a school, or even the past official rhetoric of program goals can become ammunition in a political conflict, present or future. The recording of information tends to break down the isolation of both school people and officials, exposing them to conflicts with outsiders and with each other. And it forces school officials and people in schools to think about their shared vulnerability.

None of these phenomena requires that an evaluation be well-designed, or properly analyzed, or even that it produce any findings at all. It is the analytical enterprise itself that stimulates school people and officials to think in these distinctive ways; and this is so even when they suspect or know that the evaluation cannot, in the end, serve as a reasonable basis for "policy choices." If this is so, it may be that these side effects of evaluation, instead of being peripheral to the central action of analytic thinking, are themselves at the center of a kind of analytical thinking that we have too easily ignored.

III. Interactive analysis

The failures of the conventional model of analysis do not mean that school officials and teachers are not capable of analytic thinking--just that conventional analysis does not fit their experiences and their ways of analyzing the effectiveness of schools and classrooms. What their approaches to analysis look like can be seen, in outline, in the previous section's description of how school people actually behave during a conventional analysis. They think, and they think about teaching and learning; but their thinking is tightly bound to the ongoing episodes of give-and-take among people in schools. Because this style of thinking is based on constant and sophisticated interaction with the ongoing processes of school and classroom life, I call it interactive analysis.

Interactive analysis is defined by three characteristics: it is analysis that is distinctively local and personal; it is cumulative; and it is reflexive. These attributes of interactive analysis have grown up in response to the realities of the school setting, they are a central part of the interactions among teachers and officials, and they dramatically distinguish interactive analysis from conventional analysis. Because interactive analysis focuses on the issues and discoveries of daily life in schools, it forms the basis of school people's knowledge and thinking.

A. Analysis that is local and personal

The conventional analytic approach aims at discovering policy variables that will produce high student achievement. In its search for the educational equivalent of the "magic bullet," conventional analysts lump classrooms together for the purposes of analysis,* and treat schools and classrooms as interchangeable units whose differences can be "controlled for," statistically. By contrast, I will suggest that when people in schools examine the effectiveness of their activities for improving student achievement, they rely heavily on an approach to analysis that is strictly local and personal in its scope. In this approach, people think in terms of the particular individual circumstances in a school or classroom that make for problems and accomplishments. Local and personal analysis focuses on the specific details and issues attached to the particular people in a given classroom, on the conflicts that surround them, and on the workings of their classroom's teaching and learning processes. It is situational, rather than categorical; it looks for what is special and distinctive about a classroom, rather than what is generalizable. The reason that people in schools think along the local and personal lines of interactive analysis rather than along the lumped-together, impersonal lines of conventional analysis is that the differences among classrooms drastically constrain the kinds of observation, comparison and inference that can take place.

---

* The requirement of conventional analysis for parsimonious, functionally well-defined models means that classroom differences not explicitly measured and intended in the analysis are set aside. This treats classrooms as if they can be viewed as being alike in all respects not entered into the analysis.

Classroom differences come about through the workings of reciprocal sovereignty. In their shared struggles and their particular wants, the people in each classroom create a distinctive set of social understandings, educational approaches, problem-solving patterns and shared priorities. These hammered-out agreements, taken together, make up an educational production process that is different in detail and in gross goals and methods from those of other classrooms. And the distinctive educational production processes undermine the ability of school officials to aggregate, to compare, and to observe systematically. In these circumstances, the only comprehensible, usable information about teaching and learning is information that is specific to individual classrooms.

Dependent as they are on this highly particularized, classroom-specific information, people engaged in analysis come to focus on problems, choices and constraints that are local and personal. Since their observations—their raw data—are dominated by their exchanges with people in particular classrooms, they pay close attention to what is in front of them: the experience of teaching and learning in those special classrooms. That experience is often microscopic in the problems that it attends to: the reasons for short attention spans and odd work habits, or the most promising search strategies for discovering untried methods of overcoming individual students' learning problems.

This kind of thinking need not be less powerful than the impersonal, inclusive (and thus implicitly centralized) models of analysis that are found in textbooks; it is simply much more dispersed. And dispersed, localized analysis is capable of the sorts of adaptations and adjustments to classroom situations that are hard to take into account in conventional (centralized, distanced) analysis. It is precisely its local and personal character that allows interactive analysis to make comparisons and inferences that are tailored to the reality of highly differentiated, divergent classrooms.

B. Analysis that is cumulative

Conventional analysis aims for a reliable and authoritative description of the effectiveness of school and classroom activities; it treats those activities as if they are fixed, observable things to be measured. But people in schools are confronted by an unavoidable and radical incompleteness in their grasp of how teaching and learning takes place, because of the constantly changing nature of classroom life. Consequently, school people adopt an approach to analysis that is cumulative in its inferences about teaching and learning, by following the events of the classroom over time, gradually adding up their partial explanations for what has been happening. Cumulative analysis pays attention to the history and the unfolding development of classroom problems, and carefully takes into account the moment in time at which the classroom happens to be the object of analysis. In the interactive analysis model, the need to understand the historically unfolding process of classroom production and for frequent cross-checking of the reliability and significance of piecemeal observations means that people gradually build up an understanding of classrooms. They think cumulatively, their analysis extending over the life of the classroom.

Teaching and learning are produced as a stream of constantly readjusted actions and interactions that may never be the same twice. For good or ill, classroom life is dominated by change, adaptation and development, which are inevitable concomitants of reciprocal sovereignty. These changes leave analysts scrambling to compensate, intellectually, for the incompleteness of their observations--for the fact that today's observations of a classroom will often have little to say about how that classroom will look a month later.

The changing problems of the classroom produce a stream of unpredictable classroom events that resemble the uncertain flow of history more than the predictable progress of the assembly line. The problems of teaching and learning constantly arise in new ways, making mechanical solutions useless; even specific skills are often made irrelevant. Students and teachers are forced to alter and realign their plans, their assumptions about each other and their expectations for the future; these adaptations are not simply fine tuning, but the kind of fundamental revision of the productive life of a classroom that analysts cannot afford to miss. People in classrooms change their behavior in response to events, to each others' demands, to new information, and to unexpected opportunities and ideas; what happens in April depends on the events of December, which grew out of what took place in September. At each step of this historical process, the results of the continuing search for solutions and adaptations to specific, hard-to-solve classroom problems critically determine what will happen in the next step.

This kind of world simply doesn't mesh very well with the conventional approach to analysis. Compare this view of schooling to factory production, where one can easily observe the mechanical and repetitive work process. Or compare it to craft-based production, where a sequence of precise, skilled actions contributes to the speedy completion of a piece of work. Conventional analysis works for factory and craft production; but trying to assess classroom activities with the conventional perspective inevitably leads to the radical incompleteness that renders the analysis useless.

The sense that their information is incomplete becomes, in itself, an essential piece of information for school people: it argues for qualifications and for tentativeness in their analysis. This sense of incompleteness also means that cross-checked and confirmed information will be much more useful than raw observations, so the gradual accretion of confirming evidence over time is a crucial part of the process of analysis. Since the actions and choices that take place in a classroom today cannot be understood unless one knows what came before them, observation over time becomes a valuable analytical tool. Instead of the "snapshot" observations of conventional analysis (which assumes that policy variables are discrete, unchanging things that can be precisely observed), school people find themselves gradually piecing together a picture of each classroom's problem-solving processes that is confirmed by their experiences and observations over time. These thought processes may not answer the standard analytical questions, but they are well adapted for untangling some of the ubiquitous puzzles of schooling—why an instructional method works well in one classroom but not in another; why a teacher allocates time among students in a particular way; why compliance with a new policy is more easily obtained from some classrooms than others.

In the world of the school, information rarely has a self-evident interpretation or application. It is interlocked with the history of the classroom, with the reasons for past choices, and with the fragmentary and potentially misleading revelations of current classroom behavior. Because school people cannot avoid these things, they think their way through them by relying on the cumulative acquisition of background information, efforts to confirm ideas that were initially vague or incomplete, and the gradual clarification of events as they develop. School analysts are themselves tied to the living developments of school life; their thinking interacts with that of school people. The fact that this cumulative quality of thought extends over time and may have no clear beginning or end, may lead some to believe that it is not analysis. That would be a mistake. The cumulative nature of interactive analysis is what allows it to be just as strong, robust, and penetrating as conventional analyses aspire to be. More to the point, it is the only kind of analysis that can work for people in schools.

C. Analysis that is reflexive

Conventional analysis is self-contained; it stands back from the people and processes that are to be analyzed. It is independent, even distanced. This posture cannot be sustained in schools, because of what I have called the side effects of analysis—the fact that any school analysis triggers the active and purposive responses of the people being analyzed, who act to change the analysis. Because this fact is well known to school officials, as well as to everyone else in schools and classrooms, their thinking adapts to it and is changed by it. Analytic thinking in schools is, consequently, reflexive: it is thinking in which the responses of the people being studied become a central part of the analysis. In schools, people involved in an analysis pay very close attention to their own role in it, and to the pressures on that role; consequently, their thinking becomes reflexive and self-referential. Relfexive thinking analyzes a situation that is known to respond actively to being analyzed.

Reflexiveness arises when school people, in thinking about how to conduct an analysis, ask themselves, "How are other people's reactions to me going to alter what they do and what they tell me—so that my ability to understand what's happening is affected? And how can I take their responses into account so that I can figure out what's going on here?" Reflexiveness means that people change their behavior to take account of the foreseeable responses to the analysis—they try to make the analysis work by anticipating the criticisms and hesitations of the people whose help they need. These attempts to respond to the concerns of the people who are part of the analysis serves as a constant reminder that any analysis is an interactive affair.

Part of the phenomenon of reflexiveness grows out of the wish to avoid conflict among officials and teachers. So people reassure each other of their good faith; they explain their intentions; they describe their needs for the analysis, or their fears of what it might foreshadow. This ritual of adjustment to the other parties to the analysis gradually and subtly recalibrates the analysts' thinking to take into account the expressions and reactions of the other parties. Reflexiveness thus becomes embedded in the thinking of both school officials and teachers, with everyone becoming aware of the power that responses and interactions have in influencing thought, as well as practice.

Reflexive analysis is a hybrid of interpretations, interactions and adaptations, along with situation-based posturing. This polyglot mixture of constantly revised thinking with exchanges of conditional behaviors has the effect of inducing everyone involved in an analysis to learn about and adapt to each others' concerns. The efforts of teachers and officials to make their way through this interconnected web can lead them to build coalitions or to become narrowly defensive; to seize opportunities for cooperation or selectively to exclude the most sensitive ideas and information from exchanges with others. It is a frequent by-product of reflexive thinking that people's attention is drawn to the problems of gaining agreement on shared goals, and reducing fears of uncertain future treatment. In short, reflexive thinking places analysis in a position of equality with, rather than ostensible (but unachievable) domination of, the relationships between school officials and the people of the school.

Interactive analysis, as I have suggested, is distinguished from the conventional stereotype of what analysis ought to be by its intense exchanges with, and involvement in, the ongoing processes of school and classroom life. The distinguishing marks of the way people carry out interactive analysis in schools (its local and personal awareness, its cumulative grasp, its reflexiveness) stem from the interactions that tie school officials and teachers together. Their interactions provide the validation--the constant stream of confirming evidence--from which school people construct their understanding of each other. Interactive analysis thus differs from conventional analysis in another way: it is equally the tool of both officials and teachers; it is not a top-down imposition of hypotheses, but a many-sided struggle among the people immediately concerned with the understanding of school effectiveness. That does not mean that interactive analysis is somehow guaranteed to work; the struggle may go poorly, collapsing into defensive and insecure intuitions rather than shared insights. But--unlike the conventional model of analysis--interactive analysis is always present; even if school people do not know it, and even if they aspire to textbook-style conventional analysis, interactive analysis is what they actually practice. They practice it because they have to find ways to do analysis when conventional analysis, as we know it, won't work.

Interactive analysis tends to focus on certain kinds of ideas and puzzles. Officials acquire a great deal of knowledge about individual teachers, their responses to a range of requests and pressures, their sensitivities, and about their classroom and the classroom events that occupy them. This knowledge focuses particularly on what makes a given classroom different from other classrooms.

Teachers, in turn, acquire information about officials' goals and priorities, about how and when officials are likely to check up on teachers, and about the issues that are of particular sensitivity to officials. Teachers, already aware of the differences between their classrooms and the others in their school, also learn during an evaluation or analysis about the pressures on officials to be consistent across classrooms, to aggregate and to establish categories (of special programs, grade levels or students being served), and to coordinate disparate classrooms.

For teachers, who are far too often ignored when the uses of analysis are discussed, interactive analysis is a part of the experience that draws them to consider the impact of official policies on the production of teaching and learning in classrooms. Teachers depend on their interactions with officials when they try to understand whether policies mean what they say they mean, whether exceptions to announced policies will be allowed, or what the underlying goals of a new policy are.

All of this takes place whether or not the parties to the analysis wish it to; people in schools cannot escape the interdependencies produced by officials' and teachers' give-and-take, and these interdependencies produce distinctive kinds of knowledge as an unavoidable by-product. For teachers, this means that the conduct of an analysis or evaluation informs the object of that analysis about what the analyst's concerns are. For officials, it means that the act of analysis produces unsought knowledge about the particularized, disparate character of the objects of analysis, the people in the schools.

Bibliographical Note for Essay 5

The classic debate over the utility of conventional analysis is between Schultze, The Politics and Economics of Public Spending (Brookings Institution, 1968) and Lindblom, "The Science of 'Muddling Through,'" (Public Administration Review, Spring, 1959). My work on broader conceptions of analysis is strongly influenced by Lindblom and Cohen, Usable Knowledge (Yale University Press, 1979). For surveys of the evaluation literature in education, see Averch et al., How Effective Is Schooling? (Educational Technology Publications, 1974); Berman and McLaughlin, Federal Programs Supporting Educational Change, Vol. I: A Model of Educational Change (Rand Corporation, 1974); Weiss, ed., Evaluating Action Programs (Allyn and Bacon, 1972); and in general the series, Evaluation Studies Review Annual (Sage Publications, 1976-).

The sources cited on pages 5 through 9 of this essay are Averch et al.; Chall, Learning to Read: The Great Debate (McGraw-Hill, 1967); McLaughlin, Evaluation and Reform (Ballinger, 1975); Murnane, "Interpreting the Evidence on School Effectiveness" (Teachers College Record, Fall, 1981); and VanDeusen Lukas, "Problems in Implementing Head Start Planned Variation Models," in Rivlin and Timpane, eds., Planned Variation in Education (Brookings Institution, 1975).